

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

**特許第6691280号**  
(P6691280)

(45) 発行日 令和2年4月28日 (2020. 4. 28)

(24) 登録日 令和2年4月14日 (2020. 4. 14)

(51) Int. Cl. F I  
**G06F 16/31 (2019.01)** G O 6 F 16/31

請求項の数 15 (全 50 頁)

<p>(21) 出願番号 特願2020-8423 (P2020-8423)</p> <p>(22) 出願日 令和2年1月22日 (2020.1.22)</p> <p>審査請求日 令和2年1月22日 (2020.1.22)</p> <p>(31) 優先権主張番号 特願2019-139759 (P2019-139759)</p> <p>(32) 優先日 令和1年7月30日 (2019.7.30)</p> <p>(33) 優先権主張国・地域又は機関 日本国 (JP)</p> <p>早期審査対象出願</p>	<p>(73) 特許権者 718000266 特許庁長官 東京都千代田区霞が関3-4-3</p> <p>(74) 代理人 100099759 弁理士 青木 篤</p> <p>(74) 代理人 100123582 弁理士 三橋 真二</p> <p>(74) 代理人 100114018 弁理士 南山 知広</p> <p>(74) 代理人 100165191 弁理士 河合 章</p> <p>(74) 代理人 100133835 弁理士 河野 努</p> <p>(74) 代理人 100180806 弁理士 三浦 剛</p>
--	--

最終頁に続く

(54) 【発明の名称】 管理システム及び管理方法

(57) 【特許請求の範囲】

【請求項1】

複数の特許文献に関する情報を検索用データベースに登録するための管理システムであって、

所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、前記形式情報を取得する第1サーバと、

複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なるデータを含む書誌情報が共通のデータ形式に変換されて記憶された外部データベースから、各特許文献について、文献番号と、前記共通のデータ形式に変換された書誌情報とを取得する第2サーバと、

前記第1サーバが前記形式情報を取得した各特許文献について、前記形式情報に含まれる各データを、前記形式情報及び前記書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納し、当該各特許文献について前記第2サーバが書誌情報を取得している場合は、文献番号をキーとして、当該書誌情報に含まれる各データを前記第1テーブルの対応する位置に格納することにより、前記形式情報及び前記書誌情報を統合した第1統合情報を生成する第3サーバと、

前記データ群から、各特許文献について、文献番号と、前記内容情報とを取得し、前記

10

20

発明の内容が所定言語で記載された特許文献については当該発明の内容に基づき、前記発明の内容が前記所定言語で記載されていない特許文献については前記発明の内容が前記所定言語で記載された翻訳文に基づいて、学習モデルを用いて、各特許文献の分類情報を生成する第4サーバと、を有し、

前記第3サーバは、前記第1統合情報が完成した後に、前記第1サーバが前記形式情報を取得した特許文献について、文献番号をキーとして、前記第1統合情報に含まれる各データと、前記分類情報に含まれる各データとを、前記第1統合情報及び前記分類情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、前記第1統合情報及び前記分類情報を統合した第2統合情報を生成し、前記生成した第2統合情報を検索用データベースに登録する、

ことを特徴とする管理システム。

【請求項2】

前記第4サーバは、前記第3サーバが前記第1統合情報を生成している間に、前記分類情報を生成する、請求項1に記載の管理システム。

【請求項3】

前記第4サーバは、前記発明の内容が前記所定言語で記載されていない特許文献について、前記翻訳文を生成する、請求項1または2に記載の管理システム。

【請求項4】

前記第3サーバは、ファミリー出願に係る特許文献から生成した第2統合情報を一つの第2テーブル内で相互に対応付けて記憶し、

ファミリー出願に係る特許文献から生成された第2統合情報のインデックスとして、ファミリー出願に係る特許文献で共通のインデックスを設定する第5サーバをさらに有する、請求項1または3に記載の管理システム。

【請求項5】

前記第4サーバは、特定の出願に係る特許文献について生成した分類情報を、当該特定の出願のファミリー出願に係る特許文献の分類情報として使用する、請求項1～4の何れか一項に記載の管理システム。

【請求項6】

前記第4サーバは、各特許文献について、それぞれ次元数が異なる複数の第1特徴ベクトルをさらに生成し、

前記第2テーブルには、前記第1統合情報及び前記分類情報の各データ項目と、前記複数の第1特徴ベクトルとが所定の順序で並べられており、

前記第3サーバは、文献番号をキーとして、前記第1統合情報に含まれる各データと、前記分類情報に含まれる各データと、前記複数の第1特徴ベクトルとを前記第2テーブルの対応する位置に格納することにより、前記第1統合情報、前記分類情報及び前記複数の第1特徴ベクトルを統合して前記第2統合情報を生成する、請求項1～5の何れか一項に記載の管理システム。

【請求項7】

前記第4サーバは、それぞれ異なる複数のLSHを用いて各特許文献の第2特徴ベクトルを変換したハッシュ値を各要素とする特徴ベクトルを前記第1特徴ベクトルとして生成する、請求項6に記載の管理システム。

【請求項8】

利用者によって指定された指定データについて前記複数の第1特徴ベクトルを生成し、前記次元数の昇順又は降順に、前記指定データについて生成された第1特徴ベクトルと、各特許文献について生成された第1特徴ベクトルとを比較することにより、前記指定データに対応する特許文献を抽出する第5サーバをさらに有する、請求項6または7に記載の管理システム。

【請求項9】

利用者によって指定された指定データに対応する複数の特許文献を並べて表示するための第1表示データを生成する第5サーバと、

10

20

30

40

50

前記第 1 表示データに従って、複数の特許文献を並べて表示する端末装置と、をさらに有し、

前記第 5 サーバは、前記端末装置において利用者によって指定された特許文献又は所定時間以上連続して表示された特許文献との類似度に基づいて、前記第 1 表示データにより表示される特許文献を並べ替えた第 2 表示データを生成する、請求項 1 ~ 7 の何れか一項に記載の管理システム。

【請求項 1 0】

前記第 3 サーバは、前記第 1 サーバが前記形式情報を取得した特許文献の内、前記第 2 サーバが前記書誌情報を取得しなかった特許文献に関する情報を出力する、請求項 1 ~ 9 の何れか一項に記載の管理システム。

【請求項 1 1】

前記第 1 サーバは、前記第 2 統合情報が完成した後に、前記第 1 サーバが前記形式情報を取得した特許文献について、文献番号をキーとして、前記第 2 統合情報に含まれる各データと、前記内容情報に含まれるデータとを、前記第 2 統合情報及び前記内容情報の各データ項目が所定の順序で並べられた第 3 テーブルの対応する位置に格納することにより、前記第 2 統合情報及び前記内容情報を統合した第 3 統合情報を生成し、前記生成した第 3 統合情報を前記検索用データベースに登録する、請求項 1 ~ 1 0 の何れか一項に記載の管理システム。

【請求項 1 2】

前記第 4 サーバは、

技術分野の大分類毎に、技術分野の小分類を特定するための学習モデルを記憶し、

各特許文献の技術分野の大分類を特定し、前記特定した大分類に対応する学習モデルを用いて、各特許文献の技術分野の小分類を特定し、各特許文献の技術分野の大分類及び小分類を前記分類情報として生成する、請求項 1 ~ 1 1 の何れか一項に記載の管理システム。

【請求項 1 3】

複数の特許文献に関する情報を検索用データベースに登録するための管理システムにおける管理方法であって、

第 1 サーバが、所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、前記形式情報を取得し、

第 2 サーバが、前記データ群から、各特許文献について、文献番号を含み且つ前記形式情報と異なる書誌情報を取得し、

第 3 サーバが、前記第 1 サーバが前記形式情報を取得した各特許文献について、前記形式情報に含まれる各データを、前記形式情報及び前記書誌情報の各データ項目が所定の順序で並べられた第 1 テーブルの対応する位置に格納し、当該各特許文献について前記第 2 サーバが書誌情報を取得している場合は、文献番号をキーとして、当該書誌情報に含まれる各データを前記第 1 テーブルの対応する位置に格納することにより、前記形式情報及び前記書誌情報を統合した第 1 統合情報を生成し、

第 4 サーバが、前記データ群から、各特許文献について、文献番号と、前記内容情報とを取得し、前記発明の内容が所定言語で記載された特許文献については当該発明の内容に基づき、前記発明の内容が前記所定言語で記載されていない特許文献については前記発明の内容が前記所定言語で記載された翻訳文に基づいて、学習モデルを用いて、各特許文献の分類情報を生成することを含み、

前記第 3 サーバは、前記第 1 統合情報が完成した後に、前記第 1 サーバが前記形式情報を取得した特許文献について、文献番号をキーとして、前記第 1 統合情報に含まれる各データと、前記分類情報に含まれる各データとを、前記第 1 統合情報及び前記分類情報の各データ項目が所定の順序で並べられた第 2 テーブルの対応する位置に格納することにより

10

20

30

40

50

、前記第1統合情報及び前記分類情報を統合した第2統合情報を生成し、前記生成した第2統合情報を検索用データベースに登録する、  
ことを特徴とする管理方法。

【請求項14】

複数の特許文献に関する情報を検索用データベースに登録するための管理システムであって、

所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、前記形式情報を取得する第1サーバと、

複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なるデータを含む書誌情報が共通のデータ形式に変換されて記憶された外部データベースから、各特許文献について、文献番号と、前記共通のデータ形式に変換された書誌情報とを取得する第2サーバと、

前記第1サーバが前記形式情報を取得した各特許文献について、前記形式情報に含まれる各データを、前記形式情報及び前記書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納することにより、前記形式情報及び前記書誌情報を統合した第1統合情報を生成する第3サーバと、

前記データ群から、各特許文献について文献番号と前記内容情報とを取得し、前記内容情報に基づいて学習モデルを用いて各特許文献の検索情報を生成する第4サーバと、を有し、

前記第3サーバは、前記第1統合情報が完成した後に、前記第1サーバが前記形式情報を取得した特許文献について、文献番号をキーとして、前記第1統合情報に含まれる各データと、前記第1統合情報が完成する前に生成された前記検索情報に含まれる各データを、前記第1統合情報及び前記検索情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、前記第1統合情報と、前記第1統合情報が完成する前に生成された前記検索情報とを統合した第2統合情報を生成し、前記生成した第2統合情報を検索用データベースに登録する、

ことを特徴とする管理システム。

【請求項15】

複数の特許文献に関する情報を検索用データベースに登録するための管理システムにおける管理方法であって、

第1サーバが、所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、前記形式情報を取得し、

第2サーバが、複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なるデータを含む書誌情報が共通のデータ形式に変換されて記憶された外部データベースから、各特許文献について、文献番号と、前記共通のデータ形式に変換された書誌情報とを取得し、

第3サーバが、前記第1サーバが前記形式情報を取得した各特許文献について、前記形式情報に含まれる各データを、前記形式情報及び前記書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納することにより、前記形式情報及び前記書誌情報を統合した第1統合情報を生成し、

第4サーバが、前記データ群から、各特許文献について文献番号と前記内容情報とを取得し、前記内容情報に基づいて学習モデルを用いて各特許文献の検索情報を生成することを含み、

前記第3サーバは、前記第1統合情報が完成した後に、前記第1サーバが前記形式情報

を取得した特許文献について、文献番号をキーとして、前記第1統合情報に含まれる各データと、前記第1統合情報が完成する前に生成された前記検索情報に含まれる各データを、前記第1統合情報及び前記検索情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、前記第1統合情報と、前記第1統合情報が完成する前に生成された前記検索情報とを統合した第2統合情報を生成し、前記生成した第2統合情報を検索用データベースに登録する、

ことを特徴とする管理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、特許文献を管理する管理システム及び管理方法に関する。

【背景技術】

【0002】

特許庁では、特許出願の審査時等に先行技術を調査するために、過去に出願された特許文献の検索用データベースが利用されている。この検索用データベースには、世界各国で出願された特許文献に関する情報が記憶されており、管理システムにより、最新の特許文献に関する情報が適宜登録されるように更新される。しかしながら、世界各国において常時膨大な量の特許が出願されており、管理システムはデータベースの更新に膨大な時間を要している。

【0003】

例えば、特許文献1には、図書館の蔵書を管理するシステムにおいて、ユーザの趣向に応じてお勧めの図書を提示するレコメンドシステムが開示されている。このレコメンドシステムは、高評価図書リスト及び低貸出頻度図書リストを生成し、高評価図書リスト及び低貸出頻度図書リストのいずれにも図書IDが含まれている図書のリストを抽出することにより、提示対象の図書のリストを生成する。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2015-022507号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

特許文献に関する情報を検索用データベースに登録するための管理システムでは、特許文献に関する情報を効率良く管理することが求められている。

【0006】

管理システム及び管理方法の目的は、特許文献に関する情報を効率良く管理することにある。

【課題を解決するための手段】

【0007】

実施形態に係る管理システムは、複数の特許文献に関する情報を検索用データベースに登録するための管理システムであって、所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、形式情報を取得する第1サーバと、複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なるデータを含む書誌情報が共通のデータ形式に変換されて記憶された外部データベースから、各特許文献について、文献番号と、共通のデータ形式に変換された書誌情報とを取得する第2サーバと、第1サーバが形式情報を取得した各特許文献について、形式情報に含まれる各データを、形式情報及び書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納し

10

20

30

40

50

、当該各特許文献について第2サーバが書誌情報を取得している場合は、文献番号をキーとして、当該書誌情報に含まれる各データを第1テーブルの対応する位置に格納することにより、形式情報及び書誌情報を統合した第1統合情報を生成する第3サーバと、データ群から、各特許文献について、文献番号と、内容情報とを取得し、発明の内容が所定言語で記載された特許文献については当該発明の内容に基づき、発明の内容が所定言語で記載されていない特許文献については発明の内容が所定言語に翻訳された翻訳文に基づいて、学習モデルを用いて、各特許文献の分類情報を生成する第4サーバと、を有し、第3サーバは、第1統合情報が完成した後に、第1サーバが形式情報を取得した特許文献について、文献番号をキーとして、第1統合情報に含まれる各データと、分類情報に含まれる各データとを、第1統合情報及び分類情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、第1統合情報及び分類情報を統合した第2統合情報を生成し、生成した第2統合情報を検索用データベースに登録する。

【0008】

実施形態に係る管理システムにおいて、第4サーバは、第3サーバが第1統合情報を生成している間に、分類情報を生成することが好ましい。

【0009】

実施形態に係る管理システムにおいて、第4サーバは、発明の内容が所定言語で記載されていない特許文献について、翻訳文を生成することが好ましい。

【0010】

実施形態に係る管理システムにおいて、第3サーバは、ファミリー出願に係る特許文献から生成した第2統合情報を一つの第2テーブル内で相互に対応付けて記憶し、ファミリー出願に係る特許文献から生成された第2統合情報のインデックスとして、ファミリー出願に係る特許文献で共通のインデックスを設定する第5サーバをさらに有することが好ましい。

【0011】

実施形態に係る管理システムにおいて、第3サーバは、学習モデルを用いて、各特許文献の技術分野又は翻訳データを検索情報として生成することが好ましい。

【0012】

実施形態に係る管理システムにおいて、第4サーバは、特定の出願に係る特許文献について生成した分類情報を、当該特定の出願のファミリー出願に係る特許文献の分類情報として使用することが好ましい。

【0013】

実施形態に係る管理システムにおいて、第4サーバは、各特許文献について、それぞれ次元数が異なる複数の第1特徴ベクトルをさらに生成し、第2テーブルには、第1統合情報及び分類情報の各データ項目と、複数の第1特徴ベクトルとが所定の順序で並べられており、第3サーバは、文献番号をキーとして、第1統合情報に含まれる各データと、分類情報に含まれる各データと、複数の第1特徴ベクトルとを第2テーブルの対応する位置に格納することにより、第1統合情報、分類情報及び複数の第1特徴ベクトルを統合して第2統合情報を生成することが好ましい。

【0014】

実施形態に係る管理システムにおいて、第4サーバは、それぞれ異なる複数のLSHを用いて各特許文献の第2特徴ベクトルを変換したハッシュ値を各要素とする特徴ベクトルを第1特徴ベクトルとして生成することが好ましい。

【0015】

実施形態に係る管理システムにおいて、利用者によって指定された指定データについて複数の第1特徴ベクトルを生成し、次元数の昇順又は降順に、指定データについて生成された第1特徴ベクトルと、各特許文献について生成された第1特徴ベクトルとを比較することにより、指定データに対応する特許文献を抽出する第5サーバをさらに有することが好ましい。

【0016】

10

20

30

40

50

実施形態に係る管理システムにおいて、利用者によって指定された指定データに対応する複数の特許文献を並べて表示するための第1表示データを生成する第5サーバと、第1表示データに従って、複数の特許文献を並べて表示する端末装置と、をさらに有し、第5サーバは、端末装置において利用者によって指定された特許文献又は所定時間以上連続して表示された特許文献との類似度に基づいて、第1表示データにより表示される特許文献を並べ替えた第2表示データを生成することが好ましい。

【0017】

実施形態に係る管理システムにおいて、第3サーバは、第1サーバが形式情報を取得した特許文献の内、第2サーバが書誌情報を取得しなかった特許文献に関する情報を出力することが好ましい。

【0018】

実施形態に係る管理システムにおいて、第1サーバは、第2統合情報が完成した後に、第1サーバが形式情報を取得した特許文献について、文献番号をキーとして、第2統合情報に含まれる各データと、内容情報に含まれるデータとを、第2統合情報及び内容情報の各データ項目が所定の順序で並べられた第3テーブルの対応する位置に格納することにより、第2統合情報及び内容情報を統合した第3統合情報を生成し、生成した第3統合情報を検索用データベースに登録することが好ましい。

【0019】

実施形態に係る管理システムにおいて、第4サーバは、技術分野の大分類毎に、技術分野の小分類を特定するための学習モデルを記憶し、各特許文献の技術分野の大分類を特定し、特定した大分類に対応する学習モデルを用いて、各特許文献の技術分野の小分類を特定し、各特許文献の技術分野の大分類及び小分類を分類情報として生成することが好ましい。

【0020】

実施形態に係る管理方法は、複数の特許文献に関する情報を検索用データベースに登録するための管理システムにおける管理方法であって、第1サーバが、所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、形式情報を取得し、第2サーバが、データ群から、各特許文献について、文献番号を含み且つ形式情報と異なる書誌情報を取得し、第3サーバが、第1サーバが形式情報を取得した各特許文献について、形式情報に含まれる各データを、形式情報及び書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納し、当該各特許文献について第2サーバが書誌情報を取得している場合は、文献番号をキーとして、当該書誌情報に含まれる各データを第1テーブルの対応する位置に格納することにより、形式情報及び書誌情報を統合した第1統合情報を生成し、第4サーバが、データ群から、各特許文献について、文献番号と、内容情報とを取得し、発明の内容が所定言語で記載された特許文献については当該発明の内容に基づき、発明の内容が所定言語で記載されていない特許文献については発明の内容が所定言語で記載された翻訳文に基づいて、学習モデルを用いて、各特許文献の分類情報を生成することを含み、第3サーバは、第1統合情報が完成した後に、第1サーバが形式情報を取得した特許文献について、文献番号をキーとして、第1統合情報に含まれる各データと、分類情報に含まれる各データとを、第1統合情報及び分類情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、第1統合情報及び分類情報を統合した第2統合情報を生成し、生成した第2統合情報を検索用データベースに登録する。

【0021】

実施形態に係る管理システムは、複数の特許文献に関する情報を検索用データベースに登録するための管理システムであって、所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文

献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、形式情報を取得する第1サーバと、複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なるデータを含む書誌情報が共通のデータ形式に変換されて記憶された外部データベースから、各特許文献について、文献番号と、共通のデータ形式に変換された書誌情報とを取得する第2サーバと、第1サーバが形式情報を取得した各特許文献について、形式情報に含まれる各データを、形式情報及び書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納することにより、形式情報及び書誌情報を統合した第1統合情報を生成する第3サーバと、データ群から、各特許文献について文献番号と内容情報とを取得し、内容情報に基づいて学習モデルを用いて各特許文献の検索情報を生成する第4サーバと、を有し、第3サーバは、第1統合情報が完成した後に、第1サーバが形式情報を取得した特許文献について、文献番号をキーとして、第1統合情報に含まれる各データと、第1統合情報が完成する前に生成された検索情報に含まれる各データとを、第1統合情報及び検索情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、第1統合情報と、第1統合情報が完成する前に生成された検索情報とを統合した第2統合情報を生成し、生成した第2統合情報を検索用データベースに登録する。

【0022】

実施形態に係る管理方法は、複数の特許文献に関する情報を検索用データベースに登録するための管理システムにおける管理方法であって、第1サーバが、所定期間に、自国の特許庁及び複数の他国の特許庁が有するデータベースからそれぞれ収集した、国毎にデータ形式が異なる複数の特許文献について、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む形式情報と、少なくとも各特許文献の発明の内容を含む内容情報とが含まれるデータ群から、各特許文献について、形式情報を取得し、第2サーバが、複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なるデータを含む書誌情報が共通のデータ形式に変換されて記憶された外部データベースから、各特許文献について、文献番号と、共通のデータ形式に変換された書誌情報とを取得し、第3サーバが、第1サーバが形式情報を取得した各特許文献について、形式情報に含まれる各データを、形式情報及び書誌情報の各データ項目が所定の順序で並べられた第1テーブルの対応する位置に格納することにより、形式情報及び書誌情報を統合した第1統合情報を生成し、第4サーバが、データ群から、各特許文献について文献番号と内容情報とを取得し、内容情報に基づいて学習モデルを用いて各特許文献の検索情報を生成することを含み、第3サーバは、第1統合情報が完成した後に、第1サーバが形式情報を取得した特許文献について、文献番号をキーとして、第1統合情報に含まれる各データと、第1統合情報が完成する前に生成された検索情報に含まれる各データとを、第1統合情報及び検索情報の各データ項目が所定の順序で並べられた第2テーブルの対応する位置に格納することにより、第1統合情報と、第1統合情報が完成する前に生成された検索情報とを統合した第2統合情報を生成し、生成した第2統合情報を検索用データベースに登録する。

【発明の効果】

【0023】

管理システム及び管理方法は、特許文献に関する情報を効率良く管理することができる。

【0024】

本発明の目的及び効果は、特に請求項において指摘される構成要素及び組み合わせを用いることによって認識され且つ得られるだろう。前述の一般的な説明及び後述の詳細な説明の両方は、例示的及び説明的なものであり、特許請求の範囲に記載されている本発明を制限するものではない。

【図面の簡単な説明】

【0025】

【図1】管理システム1による処理の一例を説明するための模式図である。



- 【図 2】管理システム 1 による処理の一例を説明するための模式図である。
- 【図 3】管理システム 1 による処理の一例を説明するための模式図である。
- 【図 4】実施形態に係る管理システム 1 の概略構成を示す図である。
- 【図 5】照会サーバ 100 の概略構成を示す図である。
- 【図 6】書誌サーバ 200 の概略構成を示す図である。
- 【図 7】管理サーバ 300 の概略構成を示す図である。
- 【図 8】AIサーバ 400 の概略構成を示す図である。
- 【図 9】検索サーバ 500 の概略構成を示す図である。
- 【図 10】管理テーブル 511 のデータ構造の一例を示す模式図である。
- 【図 11】管理テーブル 511 のデータ構造の一例を示す模式図である。
- 【図 12】管理テーブル 511 のデータ構造の一例を示す模式図である。
- 【図 13】管理テーブル 511 のデータ構造の一例を示す模式図である。
- 【図 14】更新処理に係る動作シーケンスの一例を示す。
- 【図 15】形式情報のデータ構造の一例を示す模式図である。
- 【図 16】書誌情報のデータ構造の一例を示す模式図である。
- 【図 17】第 1 統合情報のデータ構造の一例を示す模式図である。
- 【図 18】検索情報のデータ構造の一例を示す模式図である。
- 【図 19】第 2 統合情報のデータ構造の一例を示す模式図である。
- 【図 20】第 3 統合情報のデータ構造の一例を示す模式図である。
- 【図 21】検索情報生成処理の動作の例を示すフローチャートである。
- 【図 22】各処理の実行タイミングについて説明するための模式図である。
- 【図 23】出力処理に係る動作シーケンスの一例を示す。
- 【図 24】検索画面 2400 の一例を示す模式図である。
- 【図 25】検索処理の動作の例を示すフローチャートである。
- 【図 26】更新処理の動作の例を示すフローチャートである。
- 【図 27】他の管理システム 2 による処理の一例を説明するための模式図である。
- 【発明を実施するための形態】

【0026】

以下、実施形態の一側面に係る管理システム及び管理方法について図を参照しつつ説明する。但し、本発明の技術的範囲はそれらの実施の形態に限定されず、特許請求の範囲に記載された発明とその均等物に及ぶ点に留意されたい。

【0027】

図 1～3 は、実施形態に係る管理システム 1 による処理の一例を説明するための模式図である。

【0028】

管理システム 1 は、複数の特許文献に関する情報を検索用データベースに登録するためのシステムである。各特許文献は、特許出願又は特許登録に係る文献であり、所定の文献形式に基づいて構成され且つ書誌情報及び内容情報を含んでいる。所定の文献形式は、その特許文献が格納される格納アドレス、その特許文献に付与されるファイル名又はその特許文献が記載された言語等である。各特許文献のデータ形式（フォーマット）は、特許出願又は特許登録が行われる国毎に異なる。各特許文献は、各特許文献の文献形式に応じた形式情報を用いて管理される。形式情報は、少なくとも文献番号を含み且つ特許文献を一意に識別可能なデータを含む。書誌情報は、その特許文献に記載された出願番号、発行日、出願日、優先権情報等の書誌的事項に関する情報であり、各特許文献を特定するために使用される。内容情報は、その特許文献に記載された内容であり、特許請求の範囲、明細書、図面及び要約書等である。即ち、内容情報は、少なくとも各特許文献の発明の内容を含む。形式情報は、各国で共通のデータ形式を有しており、書誌情報及び内容情報は、国毎にデータ形式が異なるデータを含んでいる。

【0029】

図 1～3 に示すように、管理システム 1 は、照会サーバ 100 と、書誌サーバ 200 と

、管理サーバ300と、AI (Artificial Intelligence) サーバ400と、検索サーバ500とを有する。照会サーバ100は第1サーバの一例であり、書誌サーバ200は第2サーバの一例であり、管理サーバ300は第3サーバの一例であり、AIサーバ400は第4サーバの一例である。検索サーバ500は、検索用データベース600を有している。検索用データベース600には、過去に収集された既存の特許文献に関する情報が記憶されている。管理システム1には、情報処理装置15と、自国の特許庁が有する第1データベース16と、複数の他国の特許庁が有する第2データベース17と、外部データベース18とが通信接続される。各特許文献は、各国特許庁に出願又は登録され、第1データベース16、第2データベース17及び外部データベース18に記憶されている。

#### 【0030】

図1に示すように、情報処理装置15は、第1データベース16及び第2データベース17から所定期間に出願又は登録された新規の特許文献を収集し、各特許文献の形式情報及び内容情報をデータ群151として記憶する。一方、外部データベース18は、複数の国の特許庁に出願又は登録された特許文献を任意のタイミングで収集し、収集した各特許文献の書誌情報を各国で共通のデータ形式に変換して、各特許文献の文献番号と関連付けて記憶する。

#### 【0031】

図2に示すように、照会サーバ100は、情報処理装置15のデータ群151から、情報処理装置15が所定期間に収集した各特許文献について、形式情報を取得する。書誌サーバ200は、外部データベース18から、情報処理装置15が所定期間に収集した各特許文献について、文献番号と、共通のデータ形式に変換された書誌情報とを取得する。但し、外部データベース18は、独自の意向に従って任意の複数の国の特許庁に出願又は登録された特許文献を任意のタイミングで収集するため、情報処理装置15が所定期間に収集した特許文献の一部について書誌情報を記憶していない可能性がある。一方、情報処理装置15は、自国の特許庁の意向に従って、必要十分な特許文献の形式情報を必ず記憶している。即ち、情報処理装置15と外部データベース18とは、各特許文献について取得している情報と、各情報を取得するタイミングとが異なっている。

#### 【0032】

管理サーバ300は、各特許文献について、形式情報及び書誌情報の各データ項目が所定の順序で並べられた第1テーブルを記憶する。管理サーバ300は、照会サーバ100が形式情報を取得した各特許文献について、照会サーバ100から形式情報を取得し、形式情報に含まれる各データを第1テーブルの対応する位置に格納する。また、管理サーバ300は、照会サーバ100が形式情報を取得した各特許文献について、書誌サーバ200が書誌情報を取得している場合は、文献番号をキーとして書誌情報に含まれる各データを第1テーブルの対応する位置に格納する。一方、管理サーバ300は、各特許文献について書誌サーバ200が書誌情報を取得していない場合は、第1テーブルの書誌情報に対応する位置に空白を設定する。これにより、管理サーバ300は、各特許文献について、形式情報及び書誌情報を統合した第1統合情報を生成する。なお、書誌サーバ200が特許文献自体を収集して書誌情報を生成する場合、書誌サーバ200は必ず書誌情報を取得できるため、第1テーブルの書誌情報に対応する位置に空白を設定する処理は省略されてもよい。

#### 【0033】

図3に示すように、AIサーバ400は、情報処理装置15のデータ群151から、情報処理装置15が所定期間に収集した各特許文献について、文献番号と、内容情報とを取得する。AIサーバ400は、管理サーバ300が、第1統合情報を生成している間に、各特許文献の書誌情報又は内容情報に基づいて、各特許文献には記載されず且つ検索に利用される検索情報を生成する。検索情報は、外国特許文献の機械翻訳文、特許文献の技術分野の分類(特許分類)を示す分類情報、特許文献中に開示された発明の技術的特徴を表すキーワード、特許文献中の図面のメタデータ、特許文献の特徴を示す特徴ベクトル等である。特に、AIサーバ400は、発明の内容が自国の言語、即ち管理システム1が適用

10

20

30

40

50

される検索システムを提供する特許庁における使用言語で記載されていない特許文献 1 1 について、発明の内容が自国の言語に翻訳された翻訳文を生成する。A I サーバ 4 0 0 は、発明の内容が自国の言語で記載された特許文献 0 1 についてはその発明の内容に基づき、発明の内容が自国の言語で記載されていない特許文献 1 1 については生成した翻訳文に基づいて、学習モデルを用いて各特許文献の分類情報を生成する。

#### 【 0 0 3 4 】

管理サーバ 3 0 0 は、各特許文献について、第 1 統合情報及び検索情報の各データ項目が所定の順序で並べられた第 2 テーブルを記憶する。管理サーバ 3 0 0 は、第 1 統合情報が完成した後に、照会サーバ 1 0 0 が形式情報を取得した各特許文献について、文献番号をキーとして、第 1 統合情報に含まれる各データと、検索情報に含まれる各データとを第 2 テーブルの対応する位置に格納する。これにより、管理サーバ 3 0 0 は、第 1 統合情報及び検索情報を統合した第 2 統合情報を生成する。管理サーバ 3 0 0 は、生成した第 2 統合情報を、検索サーバ 5 0 0 の検索用データベース 6 0 0 に登録する。

#### 【 0 0 3 5 】

照会サーバ 1 0 0 及び書誌サーバ 2 0 0 は、情報処理装置 1 5 が所定期間に収集した各特許文献に対して短期間で、情報処理装置 1 5 のデータ群 1 5 1 から形式情報を取得し、外部データベース 1 8 から書誌情報を取得することができる。しかしながら、A I サーバ 4 0 0 は、各特許文献には記載されていない検索情報を生成するため、検索情報の生成には長時間を要する。管理サーバ 3 0 0 は、検索情報の完成を待たずに、まず、別個の装置から取得した形式情報と書誌情報を統合して第 1 統合情報を生成し、その後、検索情報を統合して第 2 統合情報を生成するため、第 2 統合情報を効率良く短時間に生成することができる。また、管理サーバ 3 0 0 は、情報処理装置 1 5 が所定期間に収集した複数の特許文献について、検索用データベース 6 0 0 に登録するために必要な第 2 統合情報をまとめて生成する。これにより、管理サーバ 3 0 0 は、情報処理装置 1 5 が所定期間に収集した複数の特許文献に関する情報を検索用データベース 6 0 0 に一括して登録することができ、検索用データベース 6 0 0 を効率良く短時間に更新することができる。したがって、管理システム 1 は、特許文献に関する情報を効率良く管理することが可能となる。

#### 【 0 0 3 6 】

図 4 は、実施形態に係る管理システム 1 の概略構成を示す図である。

#### 【 0 0 3 7 】

管理システム 1 は、特許文献に含まれる各情報を管理し、技術分類又はキーワード等の各情報を用いて特許文献を検索するためのサービスを利用者に提供する。管理システム 1 は、例えば、特許情報プラットフォーム ( J - P l a t P a t ( 登録商標 ) )、外国特許情報サービス ( F O P I S E R )、E s p a c e n e t ( 登録商標 )、P A T E N T S C O P E ( 登録商標 ) 等の検索システムに適用される。なお、管理システム 1 は、他の各国特許庁が提供する特許文献の検索システムに適用されてもよい。以下では、管理システム 1 が適用される検索システムを提供する特許庁を対象特許庁と称する場合がある。

#### 【 0 0 3 8 】

特許情報プラットフォームは、日本、米国、欧州特許庁 ( E P O )、イギリス、ドイツ、フランス、スイス、世界知的所有権機関 ( W I P O )、カナダ、韓国、中国等の特許・実用新案の各種公報及び C S D B ( Computer Software Data Base ) の各種文献を記憶するデータベースを有する。外国特許情報サービスは、ロシア、台湾、オーストラリア、シンガポール、ベトナム、タイ等の特許・実用新案の各種公報を記憶するデータベースを有する。E s p a c e n e t は、欧州特許庁が提供する 1 0 0 か国以上の特許公報等を記憶するデータベースを有する。P A T E N T S C O P E は、公開済みの P C T 国際出願 3 4 3 万件を含む、7 1 9 6 万件の特許文献を記憶するデータベースを有する。

#### 【 0 0 3 9 】

図 4 に示すように、管理システム 1 は、照会サーバ 1 0 0、書誌サーバ 2 0 0、管理サーバ 3 0 0、A I サーバ 4 0 0 及び検索サーバ 5 0 0 等を有する。管理システム 1 は、さらに、複数の端末装置 1 0、ゲートウェイサーバ 1 1、U I ( User Interface ) サーバ 1

2、バックアップサーバ13、ログ管理サーバ14、情報処理装置15、第1データベース16、複数の第2データベース17及び外部データベース18等を有する。複数の端末装置10、ゲートウェイサーバ11及びUIサーバ12は、それぞれ第1ネットワーク20に通信接続する。照会サーバ100、書誌サーバ200、管理サーバ300、AIサーバ400、検索サーバ500、ゲートウェイサーバ11、UIサーバ12、バックアップサーバ13、ログ管理サーバ14、情報処理装置15及び外部データベース18は、それぞれ第2ネットワーク21に通信接続する。情報処理装置15、第1データベース16及び第2データベース17は、それぞれ第3ネットワーク22に通信接続する。第1ネットワーク20、第2ネットワーク21及び第3ネットワーク22は、端末装置10の利用者の事業所及び各サーバの設置場所に設けられたローカルエリアネットワーク又はクラウドネットワーク等である。管理システム1は、各サーバを、それぞれ複数有してもよい。各サーバは、物理サーバである。なお、各サーバは、統合されていてもよく、各サーバは、仮想サーバでもよい。

#### 【0040】

各端末装置10は、パーソナルコンピュータ、タブレット端末、スマートフォン等であり、特許文献を検索する利用者により使用される。各端末装置10は、表示装置、入力装置、記憶装置、メモリ、CPU及び通信インタフェース回路等を有する。

#### 【0041】

ゲートウェイサーバ11は、検索サーバ500と各端末装置10の間の通信を中継するサーバであり、各端末装置10からの指示に従って検索サーバ500に特許文献の検索を指示し、検索サーバ500から検索結果を受信して各端末装置10に送信する。

#### 【0042】

UIサーバ12は、特許文献を検索するための検索画面を提供するサーバであり、各端末装置10からの指示に従って、検索画面を表示するための表示データを各端末装置10に送信する。

#### 【0043】

バックアップサーバ13は、管理サーバ300及び検索サーバ500に格納された各情報を定期的にバックアップするサーバである。バックアップのタイミングは、データベース16に記憶されたデータの更新開始前が望ましいが、それに限定されない。各サーバに障害が発生した場合、管理システム1は、バックアップサーバ13に格納された情報に基づいて、各サーバを復旧させることができ、サービスの継続性を向上させることが可能となる。

#### 【0044】

ログ管理サーバ14は、管理サーバ300からの指示に従って、照会サーバ100及び検索サーバ500等の稼働状態又は第1データベース16の記憶データ等を監視し、サーバ管理者又は利用者に監視結果を通知するサーバである。ログ管理サーバ14は、サーバの稼働状態をリアルタイムに監視し、グラフ又は表で図示した画像を用いて、サーバ管理者に通知する。これにより、サーバ管理者は、異常発生時にその異常を早期に回復させることができる。また、ログ管理サーバ14は、第1データベース16におけるデータ更新の前後に、第1データベース16の記憶データを確認し、サーバ管理者及び利用者に通知する。これにより、サーバ管理者及び利用者は、第1データベース16の記憶装置の空き容量不足によりデータを更新できなくなることを未然に防ぐことができる。

#### 【0045】

また、ログ管理サーバ14は、定期的に、照会サーバ100、管理サーバ300又は検索サーバ500から、新たに追加された特許文献に関する情報を収集し、収集した情報を発行年毎又は発行機関毎に集計する。ログ管理サーバ14は、集計した情報をグラフ又は表で図示した画像を用いて、システム管理者又は利用者に通知する。これにより、システム管理者又は利用者は、発行年毎又は発行機関毎の特許文献の件数の分布等を把握することができ、ログ管理サーバ14は、利用者の利便性を向上させることが可能となる。

#### 【0046】

10

20

30

40

50

情報処理装置 15 は、定期的に、第 1 データベース 16 及び第 2 データベース 17 から、新たに出願又は登録された特許文献を収集し、照会サーバ 100 及び AI サーバ 400 に配信する。情報処理装置 15 は、所定期間毎に、第 3 ネットワーク 22 を介して第 1 データベース 16 及び第 2 データベース 17 に、その所定期間に新たに出願、登録又は更新された特許文献の取得要求信号を送信する。情報処理装置 15 は、第 1 データベース 16 及び第 2 データベース 17 から各特許文献を受信した場合、受信した各特許文献の形式情報及び内容情報をデータ群 151 として記憶する。即ち、データ群 151 には、情報処理装置 15 が所定期間に第 1 データベース 16 及び第 2 データベース 17 からそれぞれ収集した複数の特許文献についての形式情報及び内容情報が含まれる。また、情報処理装置 15 は、受信した各特許文献の形式情報及び内容情報を、第 2 ネットワーク 21 を介して照会サーバ 100 及び AI サーバ 400 のそれぞれに送信する。 10

#### 【0047】

また、情報処理装置 15 は、取得した各特許文献にファミリー出願が存在するか否かを判定し、ファミリー出願が存在する場合、そのファミリー出願に係る特許文献を第 1 データベース 16 及び第 2 データベース 17 から取得する。また、情報処理装置 15 は、各特許文献とともに、各特許文献が記載された言語を示す情報、及び、各特許文献の最新の更新日も取得する。情報処理装置 15 は、取得した各特許文献の形式情報及び内容情報をデータ群 151 として記憶するとともに、照会サーバ 100 及び AI サーバ 400 に送信する。なお、情報処理装置 15 は第 2 ネットワーク 21 と通信接続されずに、情報処理装置 15 の管理者が、USB (Universal Serial Bus) メモリ等を利用して、受信した各特許 20  
文献の形式情報及び内容情報を照会サーバ 100 及び AI サーバ 400 に複写してもよい。

#### 【0048】

第 1 データベース 16 は、自国の特許庁 (対象特許庁) が有するデータベースであり、例えば日本国特許庁 (JPO) において出願又は登録された特許文献を記憶する。第 1 データベース 16 の数は、1 つに限定されず、複数でもよい。

#### 【0049】

複数の第 2 データベース 17 は、それぞれ複数の他国の特許庁 (対象特許庁以外の特許庁) が有するデータベースである。各第 2 データベース 17 は、例えば米国特許商標庁 (USPTO)、欧州特許庁 (EPO)、世界知的所有権機関 (WIPO)、中国特許庁 (SIPO)、ドイツ特許商標庁 (DPMA)、韓国特許庁 (KIPO) 等の各国特許庁において出願又は登録された特許文献を記憶する。第 2 データベース 17 の数は、1 つでもよい。 30

#### 【0050】

外部データベース 18 は、第 1 データベース 16 及び第 2 データベース 17 とは異なるデータベースである。外部データベース 18 は、例えば欧州特許庁 (EPO) が管理する DocDB (Document Database) 等である。外部データベース 18 は、複数の国の特許庁に出願又は登録された特許文献を任意のタイミングで収集し、収集した各特許文献の書誌情報を各国で共通のデータ形式に変換して記憶する。即ち、外部データベース 18 には、複数の国の特許庁に出願又は登録された特許文献について、国毎にデータ形式が異なる 40  
データを含む書誌情報が共通のデータ形式に変換されて記憶される。外部データベース 18 の数は、1 つに限定されず、複数でもよい。

#### 【0051】

図 5 は、照会サーバ 100 の概略構成を示す図である。

#### 【0052】

照会サーバ 100 は、対象特許庁を含む各国特許庁が発行する各特許文献のテキストデータ及びイメージデータを取得して記憶し、各特許文献の形式に応じた形式情報を取得する。照会サーバ 100 は、利用者により文献番号が指定された照会要求信号を端末装置 10 から受信した場合、指定された文献番号に対応する特許文献のテキストデータ及びイメージデータを端末装置 10 に送信する。また、照会サーバ 100 は、特許文献の検索に用 50

いられる各種情報を検索サーバ500に提供する。図5に示すように、照会サーバ100は、第1通信装置101、第1記憶装置110及び第1制御装置120等を有する。

【0053】

第1通信装置101は、照会サーバ100が第2ネットワーク21を介して各装置と所定の通信プロトコルに従って通信を行うための通信インタフェース回路を有する。所定の通信プロトコルは、TCP/IP (Transmission Control Protocol/Internet Protocol) 等である。第1通信装置101は、第2ネットワーク21を介して各装置から受信したデータを第1制御装置120に送るとともに、第1制御装置120から受け取ったデータを、第2ネットワーク21を介して各装置に送信する。

【0054】

第1記憶装置110は、RAM (Random Access Memory)、ROM (Read Only Memory) 等のメモリ装置、ハードディスク等の固定ディスク装置、又はフレキシブルディスク、光ディスク等の可搬用の記憶装置等を有する。また、第1記憶装置110には、照会サーバ100の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、コンピュータ読み取り可能な可搬型記録媒体から公知のセットアッププログラム等を用いて第1記憶装置110にインストールされてもよい。可搬型記録媒体は、例えばCD-ROM (compact disc read only memory)、DVD-ROM (digital versatile disc read only memory) 等である。コンピュータプログラムは、所定のサーバ等からインストールされてもよい。

【0055】

第1制御装置120は、予め第1記憶装置110に記憶されているプログラムに基づいて動作するCPU (Control Processing Unit) 等のプロセッサである。なお、第1制御装置120として、DSP (digital signal processor) 等が用いられてもよい。また、第1制御装置120として、LSI (large scale integration)、ASIC (Application Specific Integrated Circuit)、FPGA (Field Programming Gate Array) 等の制御回路が用いられてもよい。第1制御装置120は、第1通信装置101及び第1記憶装置110等と接続され、これらの各部を制御するとともに、形式情報の管理制御等を行う。

【0056】

第1制御装置120は、第1記憶装置110に記憶されたコンピュータプログラムを読み取り、読み取ったコンピュータプログラムに従って動作することにより、形式情報生成部121、第3統合情報生成部122及び第3統合情報送信部123として機能する。

【0057】

図6は、書誌サーバ200の概略構成を示す図である。

【0058】

書誌サーバ200は、特許文献に記載された書誌の事項に関する書誌情報を格納する。図6に示すように、書誌サーバ200は、第2通信装置201、第2記憶装置210及び第2制御装置220等を有する。

【0059】

第2通信装置201は、第1通信装置101と同様の通信装置であり、書誌サーバ200が第2ネットワーク21を介して各装置と所定の通信プロトコルに従って通信を行うための通信インタフェース回路を有する。第2通信装置201は、第2ネットワーク21を介して各装置から受信したデータを第2制御装置220に送るとともに、第2制御装置220から受け取ったデータを、第2ネットワーク21を介して各装置に送信する。

【0060】

第2記憶装置210は、第1記憶装置110と同様の記憶装置である。また、第2記憶装置210には、書誌サーバ200の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、CD-ROM、DVD-ROM等のコンピュータ読み取り可能な可搬型記録媒体から公知のセットアッププログラム等を用いて、又は、所定のサーバ等から第2記憶装置210にインストールされて

10

20

30

40

50

もよい。

【0061】

第2制御装置220は、第1制御装置120と同様の制御装置であり、予め第2記憶装置210に記憶されているプログラムに基づいて動作する。第2制御装置220として、CPU、DSP、LSI、ASIC、FPGA等のプロセッサ又は制御回路が用いられる。第2制御装置220は、第2通信装置201及び第2記憶装置210等と接続され、これらの各部を制御するとともに、書誌情報の管理制御等を行う。

【0062】

第2制御装置220は、第2記憶装置210に記憶されたコンピュータプログラムを読み取り、読み取ったコンピュータプログラムに従って動作することにより、書誌情報生成部221として機能する。

【0063】

図7は、管理サーバ300の概略構成を示す図である。

【0064】

管理サーバ300は、照会サーバ100、書誌サーバ200、管理サーバ300、AIサーバ400及び検索サーバ500の各サーバの処理、各サーバが記憶するデータ及び各サーバ間の通信を管理する。管理サーバ300は、特許文献に関する情報を収集し、利用者が検索可能な状態になるように統合する。図7に示すように、管理サーバ300は、第3通信装置301、第3記憶装置310及び第3制御装置320等を有する。

【0065】

第3通信装置301は、第1通信装置101と同様の通信装置であり、管理サーバ300が第2ネットワーク21を介して各装置と所定の通信プロトコルに従って通信を行うための通信インタフェース回路を有する。第3通信装置301は、第2ネットワーク21を介して各装置から受信したデータを第3制御装置320に送るとともに、第3制御装置320から受け取ったデータを、第2ネットワーク21を介して各装置に送信する。

【0066】

第3記憶装置310は、第1記憶装置110と同様の記憶装置である。また、第3記憶装置310には、管理サーバ300の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、プロセッサ上で動作するソフトウェアにより実装される機能モジュールである。コンピュータプログラムは、CD-ROM、DVD-ROM等のコンピュータ読み取り可能な可搬型記録媒体から公知のセットアッププログラム等を用いて、又は、所定のサーバ等から第3記憶装置310にインストールされてもよい。

【0067】

第3制御装置320は、第1制御装置120と同様の制御装置であり、予め第3記憶装置310に記憶されているプログラムに基づいて動作する。第3制御装置320として、CPU、DSP、LSI、ASIC、FPGA等のプロセッサ又は制御回路が用いられる。第3制御装置320は、第3通信装置301及び第3記憶装置310等と接続され、これらの各部を制御するとともに、特許文献に関する各情報の管理制御等を行う。

【0068】

第3制御装置320は、第3記憶装置310に記憶されたコンピュータプログラムを読み取り、読み取ったコンピュータプログラムに従って動作することにより、第1統合情報生成部321、第2統合情報生成部322及び第2統合情報送信部323として機能する。

【0069】

図8は、AIサーバ400の概略構成を示す図である。

【0070】

AIサーバ400は、AI技術を利用して、特許文献から推定される検索情報を生成し、生成した検索情報を格納する。図8に示すように、AIサーバ400は、第4通信装置401、第4記憶装置410及び第4制御装置420等を有する。

## 【 0 0 7 1 】

第4通信装置401は、第1通信装置101と同様の通信装置であり、AIサーバ400が第2ネットワーク21を介して各装置と所定の通信プロトコルに従って通信を行うための通信インタフェース回路を有する。第4通信装置401は、第2ネットワーク21を介して各装置から受信したデータを第4制御装置420に送るとともに、第4制御装置420から受け取ったデータを、第2ネットワーク21を介して各装置に送信する。

## 【 0 0 7 2 】

第4記憶装置410は、第1記憶装置110と同様の記憶装置である。また、第4記憶装置410には、AIサーバ400の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、CD-ROM、DVD-ROM等のコンピュータ読み取り可能な可搬型記録媒体から公知のセットアッププログラム等を用いて、又は、所定のサーバ等から第4記憶装置410にインストールされてもよい。

## 【 0 0 7 3 】

第4記憶装置410には、対象特許庁における使用言語と異なる複数の言語毎に、各言語で記載された特許文献をその使用言語に翻訳するための機械翻訳エンジンである第1学習モデル411が記憶される。対象特許庁における使用言語は、所定言語の一例であり、対象特許庁が設置された国の母国語のように、対象特許庁に出願される特許文献が記載される言語等である。

## 【 0 0 7 4 】

また、第4記憶装置410には、対象特許庁により割り当てられる技術分野の大分類毎に、技術分野の小分類を特定するための第2学習モデル412が記憶される。例えば、対象特許庁が日本国特許庁である場合、技術分野の大分類はテーマコード等であり、小分類はFI又はFターム等である。即ち、対象特許庁が日本国特許庁である場合、テーマコード毎に、FI及び/又はFタームを特定するための第2学習モデル412が記憶される。なお、FI及びFタームをまとめて特定するための第2学習モデル412が記憶されてもよい。一方、対象特許庁が米国特許商標庁又は欧州特許庁である場合、技術分野の小分類はCPC (Cooperative Patent Classification、欧州米国共通特許分類) 等であり、大分類は複数のCPCの集合等である。

## 【 0 0 7 5 】

FIは、IPC (国際特許分類) を細分化した日本国特許庁独自の各特許文献の分類である。IPCは、世界知的所有権機関(WIPO)が管理する国際特許分類に関するストラスブル協定に基づいて作成された、国際的に統一されて用いられている、各特許文献の技術内容による分類である。テーマコードは、FIの各項目(約20万項目)を約2600にまとめた各テーマに割り当てられたコードであり、各特許文献の対象技術の範囲を表す。Fタームは、日本国特許庁が編纂している、各特許文献に記載された発明の技術的特徴による分類体系であり、その分類体系において用いられる分類記号である。Fタームは、IPC及びFIの特許分類体系とは異なる複数の技術的観点によって特許文献を分類する。

## 【 0 0 7 6 】

第4制御装置420は、第1制御装置120と同様の制御装置であり、予め第4記憶装置410に記憶されているプログラムに基づいて動作する。第4制御装置420として、CPU、DSP、LSI、ASIC、FPGA等のプロセッサ又は制御回路が用いられる。第4制御装置420は、第4通信装置401及び第4記憶装置410等と接続され、これらの各部を制御するとともに、検索情報の管理制御等を行う。

## 【 0 0 7 7 】

第4制御装置420は、第4記憶装置410に記憶されたコンピュータプログラムを読み取り、読み取ったコンピュータプログラムに従って動作することにより、検索情報生成部421として機能する。

## 【 0 0 7 8 】



図9は、検索サーバ500の概略構成を示す図である。

【0079】

検索サーバ500は、特許文献の検索に用いられる検索用データベース600を有し、検索用データベース600において、特許文献毎に、各特許文献の形式情報、書誌情報、検索情報及び内容情報等を一括して管理する。検索サーバ500は、ゲートウェイサーバ11を介して端末装置10から受信した利用者の指示に従って特許文献を検索し、検索結果（特許文献の文献番号等）を、ゲートウェイサーバ11を介して端末装置10に送信する。図9に示すように、検索サーバ500は、第5通信装置501、第5記憶装置510及び第5制御装置520等を有する。

【0080】

第5通信装置501は、第1通信装置101と同様の通信装置であり、検索サーバ500が第2ネットワーク21を介して各装置と所定の通信プロトコルに従って通信を行うための通信インタフェース回路を有する。第5通信装置501は、第2ネットワーク21を介して各装置から受信したデータを第5制御装置520に送るとともに、第5制御装置520から受け取ったデータを、第2ネットワーク21を介して各装置に送信する。

【0081】

第5記憶装置510は、第1記憶装置110と同様の記憶装置である。また、第5記憶装置510には、検索サーバ500の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、CD-ROM、DVD-ROM等のコンピュータ読み取り可能な可搬型記録媒体から公知のセットアッププログラム等を用いて、又は、所定のサーバ等から第5記憶装置510にインストールされてもよい。第5記憶装置510は、検索用データベースの一例である。第5記憶装置510には、データとして、管理テーブル511が記憶される。

【0082】

第5制御装置520は、第1制御装置120と同様の制御装置であり、予め第5記憶装置510に記憶されているプログラムに基づいて動作する。第5制御装置520として、CPU、DSP、LSI、ASIC、FPGA等のプロセッサ又は制御回路が用いられる。第5制御装置520は、第5通信装置501及び第5記憶装置510等と接続され、これらの各部を制御するとともに、管理テーブル511の管理制御等を行う。

【0083】

第5制御装置520は、第5記憶装置510に記憶されたコンピュータプログラムを読み取り、読み取ったコンピュータプログラムに従って動作することにより、第3統合情報記憶制御部521及び検索部522として機能する。

【0084】

図10～図13は、管理テーブル511のデータ構造の一例を示す模式図である。

【0085】

管理テーブル511には、複数の特許文献毎に、各特許文献の文献番号、基本情報、検索情報、テキストデータ、二次データ及び管理データ等が記憶される。なお、図面を簡略化するために、図10～図13には、一つのファミリー出願に対応する二つの特許文献についてのテーブルのみが示されているが、管理テーブル511には、ファミリー出願のセット毎のテーブルが含まれる。文献番号は、特許文献の公開番号又は登録番号である。

【0086】

基本情報は、各特許文献の形式情報及び書誌情報に基づく情報であり、発行機関、言語、文献種別、出願番号、発行日、出願日、テーマコード、Fターム、FI、IPC、出願人名及び発明者名等を含む。発行機関は、その特許文献を発行する特許庁である。言語は、その特許文献が記載されている言語である。文献種別は、その特許文献の種別（公開公報又は特許公報等）である。出願番号は、その特許文献に係る出願に割り当てられた出願番号である。発行日は、その特許文献が発行された日である。出願日は、その特許文献が出願された日である。出願人名は、その特許文献に係る出願の出願人の名前である。発明者名は、その特許文献に記載された発明の発明者の名前である。なお、各情報について一

10

20

30

40

50

つの特許文献に複数の値が存在する場合、各値が、カンマ等のデリミタで区切られて一つのフィールドに格納される。なお、本実施形態の各図面では、対象特許庁が日本国特許庁であり、技術分野の分類がテーマコード、Fターム、F I及びI P Cである場合の例について図示するが、技術分野の分類は、対象特許庁により割り当てられる分類に合わせて設定される。

**【 0 0 8 7 】**

検索情報は、各特許文献から推定されて生成された情報であり、テーマコード、Fターム、F I、機械翻訳文、翻訳方式、図面のメタデータ及び複数の第1特徴ベクトル等を含む。テーマコード、Fターム及びF Iは、それぞれ学習モデルを用いて推定された各特許文献のテーマコード、Fターム及びF Iである。機械翻訳文は、学習モデルを用いて各特許文献のテキストコードを対象特許庁における使用言語に翻訳した翻訳文である。翻訳方式は、各機械翻訳の方式であり、例えば統計的機械翻訳(SMT)又はニューラル機械翻訳(NMT)等である。検索情報には、一又は複数の翻訳方式毎に、その翻訳方式で翻訳された機械翻訳文が含まれる。図面のメタデータは、学習モデルを用いて推定された各特許文献の図面の特徴情報(付帯情報)である。第1特徴ベクトルは、各特許文献の特徴を示す特徴ベクトルである。また、図示されないが、検索情報には、学習モデルを用いて推定された各特許文献のキーワード等がさらに含まれてもよい。

**【 0 0 8 8 】**

テキストデータは、各特許文献に含まれるテキストデータであり、発明の名称、抄録、特許請求の範囲、詳細な説明及び全テキスト等を含む。テキストデータとして内容情報に含まれるデータが記憶される。

**【 0 0 8 9 】**

二次データは、各特許文献の形式情報、書誌情報及びテキストデータから分析されて生成された二次的(付带的)なデータであり、ファミリーID及び代表文献フラグ等を含む。ファミリーIDは、各特許文献に係るファミリー出願(同一の特許出願を基礎として各国に出願された一群の出願)に対応する特許文献を示す識別情報である。代表文献フラグは、ファミリー出願に対応する特許文献の中で最も優先度が高い特許文献を示す。

**【 0 0 9 0 】**

管理情報は、各特許文献の形式情報に基づく情報であり、更新日、格納アドレス、ファイル名、検索サーバ名及び照会サーバ名等を含む。更新日は、各特許文献の最新の更新日である。格納アドレスは、各特許文献のテキストファイルが格納されたアドレスである。ファイル名は、各特許文献のファイル名である。検索サーバ名は、各特許文献を検索する際にアクセスされるべきサーバの識別情報であり、各特許文献の管理テーブルを記憶する検索サーバの識別情報である。照会サーバ名は、各特許文献を照会する際にアクセスされるべきサーバの識別情報であり、各特許文献のテキストデータ及びイメージデータを記憶する検索サーバの識別情報である。

**【 0 0 9 1 】**

管理テーブル511では、上記した項目毎に、データ型、インデックス及び一又は複数のレコードが記憶される。データ型は、管理テーブル511に記憶されるデータの種類(文字列又は数値等)を示す。インデックスは、各レコードの索引であり、特許文献の検索に使用される。例えば、テーマコード、Fターム、F I、I P Cのインデックスとして、テーマコード、Fターム、F I、I P Cを示す文字列が設定される。また、出願人名、発明者名、翻訳文及び各テキストデータのインデックスは、英語のように空白によって区切られる言語の場合、形態素単位で設定され、日本語のように空白によって区切られない言語の場合、N-gramにより設定される。第1特徴ベクトルのインデックスとして、第1特徴ベクトルの各要素の値が設定される。他の項目のインデックスにはブランクが設定される。レコードは、データ項目の一例である。

**【 0 0 9 2 】**

管理テーブル511ではファミリー出願に対応する特許文献毎に各情報が管理され、レコードは特許文献毎に設定される。例えば、日本国出願に係る特許文献と、そのファミリ

一出願である国際出願に係る特許文献とは同一のテーブルに記憶され、各特許文献に関する情報が別個のレコードに記憶される。また、ファミリー出願に係る特許文献のインデックスとして、ファミリー出願に係る特許文献で共通のインデックスが記憶される。

【0093】

日本国特許庁に出願された特許文献にはテーマコード、Fターム及びF Iが付与され、一方、日本国特許庁以外の特許庁に出願された特許文献にはテーマコード、Fターム及びF Iが付与されない。したがって、日本国以外の特許庁に出願された特許文献の基本情報のテーマコード、Fターム及びF Iには空白が設定される。また、日本国特許庁に出願された特許文献については、A Iを用いてテーマコード、Fターム及びF Iが推定されず、検索情報のテーマコード、Fターム及びF Iには空白が設定される。同様に、日本語で出願された特許文献については、A Iを用いて機械翻訳文が生成されず、検索情報の機械翻訳文及び翻訳方式には空白が設定される。

10

【0094】

上記したように、管理システム1では、管理テーブル511において、各特許文献に関する情報が集約されて一元管理されているため、検索サーバ500は、複数のフィールドにまたがった条件が指定された検索を、簡易な検索式によって、実行することができる。例えば、検索サーバ500は、検索条件としてキーワードと図面のメタデータ等がまとめて指定された場合でも、両方の検索条件を満たす特許文献を簡易に検出することが可能となる。したがって、検索サーバ500は、ビッグデータを効率良く且つ高速に検索することができる。

20

【0095】

また、検索サーバ500は、検索条件として技術分野の分類が指定されたときに、特許文献の書誌的事項に記載された技術分野の分類と、A Iサーバ400により推定された技術分野の分類とをまとめて照合でき、効率良く且つ高速に検索することができる。同様に、検索サーバ500は、検索条件としてキーワード等が指定されたときに、特許文献の原文と、A Iサーバ400により生成された機械翻訳文とをまとめて照合でき、効率良く且つ高速に検索することができる。さらに、検索サーバ500は、検索条件としてキーワード等が指定されたときに、特許文献の原文と、A Iサーバ400により生成された第1特徴ベクトルとをまとめて照合でき、効率良く且つ高速に検索することができる。

30

【0096】

図14は、管理システム1による更新処理に係る動作シーケンスの一例を示す。

【0097】

以下に説明する動作シーケンスは、管理システム1が有する各サーバの記憶装置に予め記憶されているプログラムに基づいて、主に各サーバの制御装置により、各サーバの各要素と協働して実行される。この動作シーケンスは、一定期間（例えば1週間）毎に実行される。

【0098】

まず、情報処理装置15は、所定期間（例えば直近の1週間）に第1データベース16及び第2データベース17からそれぞれ収集した複数の特許文献について、データ群151に記憶された形式情報及び内容情報を照会サーバ100に送信する（ステップS101）。情報処理装置15は、自発的に、形式情報を照会サーバ100に送信する。なお、情報処理装置15は、照会サーバ100からの要求に従って、形式情報及び内容情報を照会サーバ100に送信してもよい。

40

【0099】

照会サーバ100の形式情報生成部121は、第1通信装置101を介して情報処理装置15から形式情報及び内容情報を受信する。これにより、形式情報生成部121は、情報処理装置15が所定期間に第1データベース16及び第2データベース17からそれぞれ収集した複数の特許文献について、データ群151から形式情報を取得する（ステップS102）。なお、形式情報生成部121は、情報処理装置15から特許文献自体を収集し、収集した特許文献毎に、各特許文献の文献形式を特定して、各特許文献の文献形式に

50

応じた形式情報を生成することにより取得してもよい。また、各特許文献から抽出される文献番号のフォーマットは、国毎に異なるため、形式情報生成部 1 2 1 は、抽出した文献番号を管理システム 1 内の共通形式のフォーマットに変換する。

【 0 1 0 0 】

図 1 5 は、形式情報のデータ構造の一例を示す模式図である。

【 0 1 0 1 】

図 1 5 に示すように、形式情報には、発行機関、文献番号、文献種別、格納アドレス、ファイル名、言語及び更新日等が含まれる。格納アドレス及びファイル名は、照会サーバ 1 0 0 の第 1 記憶装置 1 1 0 において各特許文献を格納するアドレス及びそのファイル名である。更新日は、各データベースにおいて各特許文献が更新された日である。図 1 5 に示すように、形式情報には、上記した項目毎に、データ型及び一又は複数のレコードが含まれる。形式情報生成部 1 2 1 は、各特許文献にファミリー出願が存在する場合、そのファミリー出願に係る特許文献についてデータ群 1 5 1 から形式情報を取得し、ファミリー出願に係る特許文献の形式情報を一つのテーブル内で相互に対応付けて記憶する。形式情報生成部 1 2 1 は、一つのテーブル内で、ファミリー出願に係る特許文献毎に形式情報のレコードを設定する。

【 0 1 0 2 】

一方、外部データベース 1 8 は、情報処理装置 1 5 が所定期間に第 1 データベース 1 6 及び第 2 データベース 1 7 からそれぞれ収集した複数の特許文献について、文献番号と、共通のデータ形式に変換された書誌情報とを書誌サーバ 2 0 0 に送信する（ステップ S 1 0 3 ）。外部データベース 1 8 は、書誌サーバ 2 0 0 からの要求に従って、文献番号及び書誌情報を書誌サーバ 2 0 0 に送信する。但し、外部データベース 1 8 は、複数の国の特許庁に出願又は登録された特許文献を任意のタイミングで収集するため、情報処理装置 1 5 が所定期間に収集した特許文献の一部について書誌情報を記憶していない可能性がある。そのため、外部データベース 1 8 は、情報処理装置 1 5 が所定期間に収集した特許文献の内、共通のデータ形式に変換された書誌情報を記憶している特許文献についてのみ、書誌サーバ 2 0 0 に書誌情報を送信する。

【 0 1 0 3 】

書誌サーバ 2 0 0 の書誌情報生成部 2 2 1 は、第 2 通信装置 2 0 1 を介して外部データベース 1 8 から、文献番号と、共通のデータ形式に変換された書誌情報とを受信する。これにより、書誌情報生成部 2 2 1 は、情報処理装置 1 5 が所定期間に第 1 データベース 1 6 及び第 2 データベース 1 7 からそれぞれ収集した複数の特許文献について、外部データベース 1 8 から、文献番号と、共通のデータ形式に変換された書誌情報とを取得する（ステップ S 1 0 4 ）。なお、書誌情報生成部 2 2 1 は、外部データベース 1 8 、第 1 データベース 1 6 又は第 2 データベース 1 7 から特許文献自体を収集し、収集した特許文献毎に、各特許文献に記載された書誌的事項を抽出して、書誌情報を生成してもよい。また、各特許文献から抽出される文献番号のフォーマットは、国毎に異なるため、書誌情報生成部 2 2 1 は、抽出した文献番号を管理システム 1 内の共通形式のフォーマットに変換する。

【 0 1 0 4 】

図 1 6 は、書誌情報のデータ構造の一例を示す模式図である。

【 0 1 0 5 】

図 1 6 に示すように、書誌情報には、発行機関、文献番号、文献種別、出願番号、発行日、出願日、F I、テーマコード、F ターム、I P C、出願人名、発明者名及び更新日等が含まれる。図 1 6 に示すように、書誌情報には、上記した項目毎に、データ型及び一又は複数のレコードが含まれる。書誌情報生成部 2 2 1 は、各特許文献にファミリー出願が存在する場合、そのファミリー出願に係る特許文献について外部データベース 1 8 から書誌情報を取得し、ファミリー出願に係る特許文献の書誌情報を一つのテーブル内で相互に対応付けて記憶する。書誌情報生成部 2 2 1 は、一つのテーブル内で、ファミリー出願に係る特許文献毎に書誌情報のレコードを設定する。

【 0 1 0 6 】

一方、情報処理装置 15 は、所定期間に第 1 データベース 16 及び第 2 データベース 17 からそれぞれ収集した複数の特許文献について、データ群 151 に記憶された文献番号及び内容情報を AI サーバ 400 に送信する（ステップ S105）。情報処理装置 15 は、自発的に、文献番号及び内容情報を照会サーバ 100 に送信する。なお、情報処理装置 15 は、AI サーバ 400 からの要求に従って、文献番号及び内容情報を AI サーバ 400 に送信してもよい。

#### 【0107】

AI サーバ 400 の検索情報生成部 421 は、第 4 通信装置 401 を介して情報処理装置 15 から文献番号及び内容情報を受信する。これにより、検索情報生成部 421 は、情報処理装置 15 が所定期間に第 1 データベース 16 及び第 2 データベース 17 からそれぞれ収集した複数の特許文献について、データ群 151 から文献番号及び内容情報を取得する。次に、検索情報生成部 421 は、情報処理装置 15 が所定期間に収集した各特許文献について、検索情報生成処理を実行する（ステップ S106）。検索情報生成処理において、検索情報生成部 421 は、各特許文献の内容情報に基づいて、各特許文献の検索情報を生成する。特に、検索情報生成部 421 は、学習モデルを用いて、各特許文献の分類情報、キーワード、メタデータ又は複数の第 1 特徴ベクトル等を検索情報として生成する。なお、検索情報生成部 421 は、情報処理装置 15 から特許文献自体を収集し、収集した特許文献から書誌的事項を抽出して書誌情報を生成し、生成した書誌情報にさらに基づいて検索情報を生成してもよい。検索情報生成処理の詳細については後述する。

#### 【0108】

なお、ステップ S101～S106 の各処理は、更新処理に係る動作シーケンスが実行される一定期間の開始時である第 1 タイミングに開始されて、並列して実行される。但し、ステップ S106 の処理は、ステップ S102 及び S104 の処理より長時間を要するため、ステップ S101～S104 及び後述するステップ S107～S111 の各処理と並列して実行される。即ち、検索情報生成部 421 は、管理サーバ 300 が第 1 統合情報を生成している間に検索情報を生成する。なお、AI サーバ 400 及び管理サーバ 300 は相互に独立し、検索情報生成部 421 は、管理サーバ 300 の第 1 統合情報生成部 321 が第 1 統合情報を生成する処理と並行して検索情報を生成する。第 1 統合情報生成部 321 による第 1 統合情報の生成が開始する前に、検索情報生成部 421 による検索情報の生成が完了してもよいし、検索情報生成部 421 による検索情報の生成が開始する前に、第 1 統合情報生成部 321 による第 1 統合情報の生成が完了してもよい。

#### 【0109】

一方、管理サーバ 300 の第 1 統合情報生成部 321 は、形式情報の取得を要求するための形式情報要求を、第 3 通信装置 301 を介して照会サーバ 100 に送信する（ステップ S107）。

#### 【0110】

照会サーバ 100 の形式情報生成部 121 は、第 1 通信装置 101 を介して管理サーバ 300 から形式情報要求を受信した場合、ステップ S102 で取得した形式情報を、第 1 通信装置 101 を介して管理サーバ 300 に送信する（ステップ S108）。

#### 【0111】

また、管理サーバ 300 の第 1 統合情報生成部 321 は、書誌情報の取得を要求するための書誌情報要求を、第 3 通信装置 301 を介して書誌サーバ 200 に送信する（ステップ S109）。

#### 【0112】

書誌サーバ 200 の書誌情報生成部 221 は、第 2 通信装置 201 を介して管理サーバ 300 から書誌情報要求を受信した場合、ステップ S104 で取得した書誌情報を、第 2 通信装置 201 を介して管理サーバ 300 に送信する（ステップ S110）。

#### 【0113】

なお、ステップ S107 及びステップ S109 の処理は、更新処理に係る動作シーケンスが実行される一定期間内で、第 1 タイミングより後の第 2 タイミングに実行される。第

10

20

30

40

50

2 タイミングは、第1 タイミングより、少なくとも形式情報及び書誌情報の取得が完了するための十分な期間（例えば2日間）だけ後のタイミングに設定される。

【0114】

管理サーバ300の第1統合情報生成部321は、第3通信装置301を介して照会サーバ100から形式情報を受信し、書誌サーバ200から書誌情報を受信した場合、受信した形式情報及び書誌情報を統合した第1統合情報を生成する（ステップS111）。第1統合情報生成部321は、照会サーバ100が形式情報を取得した各特許文献について、第1統合情報を生成する。

【0115】

管理サーバ300は、各特許文献について、形式情報及び書誌情報の各データ項目が所定の順序で並べられた第1テーブルを第3記憶装置310に記憶する。第1統合情報生成部321は、各特許文献について、照会サーバ100から受信した形式情報に含まれる各データを第1テーブルの対応する位置に格納する。また、第1統合情報生成部321は、各特許文献について、書誌サーバ200が外部データベース18から書誌情報を取得している場合は、書誌サーバ200から受信した文献番号をキーとして、書誌サーバ200から受信した書誌情報に含まれる各データを第1テーブルの対応する位置に格納する。一方、第1統合情報生成部321は、各特許文献について書誌サーバ200が外部データベース18から書誌情報を取得していない場合は、第1テーブルの書誌情報に対応する位置にブランクを設定する。これにより、第1統合情報生成部321は、照会サーバ100が形式情報を取得した各特許文献について、第1統合情報を生成する。

【0116】

なお、第1統合情報生成部321は、照会サーバ100が形式情報を取得した特許文献の内、書誌サーバ200が外部データベース18から書誌情報を取得しなかった特許文献に関する情報を出力してもよい。第1統合情報生成部321は、書誌サーバ200が書誌情報を取得しなかった特許文献に関する情報として、例えばその特許文献の文献番号を出力する。第1統合情報生成部321は、書誌サーバ200が書誌情報を取得しなかった特許文献に関する情報を、第3通信装置301を介してログ管理サーバ14に送信することにより出力する。なお、第1統合情報生成部321は、書誌サーバ200が書誌情報を取得しなかった特許文献に関する情報を、不図示の表示装置に表示させることにより出力してもよい。これにより、管理システム1の管理者は、第1統合情報、第2統合情報又は第3統合情報において、共通のデータ形式に変換された書誌情報が記憶されていない特許文献を判別することができ、そのような特許文献に対して個別に各情報を更新することができる。

【0117】

図17は、第1統合情報（第1テーブル）のデータ構造の一例を示す模式図である。

【0118】

図17に示すように、第1統合情報には、発行機関、文献番号、文献種別、格納アドレス、ファイル名、言語、（形式情報の）更新日、出願番号、発行日、出願日、FI、テーマコード、Fターム、IPC、出願人名、発明者名及び（書誌情報の）更新日等が含まれる。図17に示すように、第1統合情報には、上記した項目毎に、データ型及び一又は複数のレコードが含まれる。このように、第1テーブルには、形式情報及び書誌情報の各データ項目が所定の順序で並べられている。

【0119】

第1統合情報生成部321は、受信した各形式情報及び書誌情報の内、文献番号が同一である形式情報及び書誌情報の組合せを統合（マージ）することにより、第1統合情報を生成する。第1統合情報生成部321は、形式情報に含まれる発行機関、文献番号、文献種別、格納アドレス、ファイル名、言語、更新日と、書誌情報に含まれる更新日、出願番号、発行日、出願日、FI、テーマコード、Fターム、IPC、出願人名、発明者名及び更新日とを含む第1統合情報を生成する。なお、第1統合情報生成部321は、発行機関、文献番号及び文献種別を、形式情報からでなく書誌情報から抽出してもよい。第1統合

情報生成部 3 2 1 は、各特許文献にファミリー出願が存在する場合、ファミリー出願に係る各特許文献から生成した第 1 統合情報を一つの第 1 テーブル内で相互に対応付けて記憶する。第 1 統合情報生成部 3 2 1 は、一つの第 1 テーブル内で、ファミリー出願に係る特許文献毎に第 1 統合情報のレコードを設定する。

【 0 1 2 0 】

一方、A I サーバ 4 0 0 の検索情報生成部 4 2 1 は、ステップ S 1 0 6 の検索情報生成処理で生成した各検索情報を第 4 記憶装置 4 1 0 に記憶する（ステップ S 1 1 2 ）。

【 0 1 2 1 】

図 1 8 は、検索情報のデータ構造の一例を示す模式図である。

【 0 1 2 2 】

図 1 8 に示すように、検索情報には、文献番号と、テーマコードと、F タームと、F I と、一又は複数の機械翻訳文及び翻訳方式の組合せと、図面のメタデータと、複数の第 1 特徴ベクトル等とが含まれる。また、図示されないが、検索情報には、各特許文献のキーワード等がさらに含まれてもよい。図 1 8 に示すように、検索情報には、上記した項目毎に、データ型及び一又は複数のレコードが含まれる。なお、各特許文献に含まれる文献番号のフォーマットは、国毎に異なるため、検索情報生成部 4 2 1 は、各特許文献に含まれる文献番号を管理システム 1 内の共通形式のフォーマットに変換して検索情報に記憶する。また、検索情報生成部 4 2 1 は、各特許文献にファミリー出願が存在する場合、そのファミリー出願に係る特許文献についてデータ群 1 5 1 から内容情報を取得して検索情報を生成し、ファミリー出願に係る特許文献の検索情報を一つのテーブル内で相互に対応付けて記憶する。検索情報生成部 4 2 1 は、一つのテーブル内で、ファミリー出願に係る特許文献毎に検索情報を設定する。

【 0 1 2 3 】

次に、管理サーバ 3 0 0 の第 2 統合情報生成部 3 2 2 は、検索情報の取得を要求するための検索情報要求を、第 3 通信装置 3 0 1 を介して A I サーバ 4 0 0 に送信する（ステップ S 1 1 3 ）。

【 0 1 2 4 】

A I サーバ 4 0 0 の検索情報生成部 4 2 1 は、第 4 通信装置 4 0 1 を介して管理サーバ 3 0 0 から検索情報要求を受信した場合、ステップ S 1 1 2 で記憶した検索情報を、第 4 通信装置 4 0 1 を介して管理サーバ 3 0 0 に送信する（ステップ S 1 1 4 ）。

【 0 1 2 5 】

なお、ステップ S 1 1 3 の処理は、更新処理に係る動作シーケンスが実行される一定期間内で、第 2 タイミングより後の第 3 タイミングに実行される。第 3 タイミングは、第 1 タイミングより、少なくとも検索情報生成部 4 2 1 による検索情報の生成が完了するための十分な期間（例えば 3 . 5 日間）だけ後のタイミングに設定される。

【 0 1 2 6 】

管理サーバ 3 0 0 の第 2 統合情報生成部 3 2 2 は、第 3 通信装置 3 0 1 を介して A I サーバ 4 0 0 から検索情報を受信した場合、ステップ S 1 1 1 で生成した第 1 統合情報と、受信した検索情報を統合した第 2 統合情報を生成する（ステップ S 1 1 5 ）。

第 2 統合情報生成部 3 2 2 は、照会サーバ 1 0 0 が形式情報を取得した各特許文献について、第 2 統合情報を生成する。この第 2 統合情報は、検索用データベース 6 0 0 に登録を行うために使用される。

【 0 1 2 7 】

管理サーバ 3 0 0 は、各特許文献について、第 1 統合情報及び検索情報の各データ項目が所定の順序で並べられた第 2 テーブルを第 3 記憶装置 3 1 0 に記憶する。第 2 統合情報生成部 3 2 2 は、各特許文献について、文献番号をキーとして、第 1 統合情報生成部 3 2 1 が生成した第 1 統合情報に含まれる各データと、A I サーバ 4 0 0 から受信した検索情報に含まれる各データとを、第 2 テーブルの対応する位置に格納する。これにより、第 2 統合情報生成部 3 2 2 は、照会サーバ 1 0 0 が形式情報を取得した各特許文献について、第 2 統合情報を生成する。

10

20

30

40

50

## 【 0 1 2 8 】

図 1 9 は、第 2 統合情報（第 2 テーブル）のデータ構造の一例を示す模式図である。

## 【 0 1 2 9 】

図 1 9 に示すように、第 2 統合情報には、第 1 統合情報に含まれる各情報に加えて、検索情報のテーマコード、F ターム、F I、機械翻訳文、翻訳方式、図面のメタデータ及び複数の第 1 特徴ベクトル等が含まれる。図 1 9 に示すように、第 2 統合情報には、上記した項目毎に、データ型及び一又は複数のレコードが含まれる。このように、第 2 テーブルには、第 1 統合情報及び検索情報の各データ項目が所定の順序で並べられている。

## 【 0 1 3 0 】

第 2 統合情報生成部 3 2 2 は、生成した第 1 統合情報及び受信した検索情報の内、文献番号が同一である第 1 統合情報及び検索情報の組合せを統合（マージ）することにより、第 2 統合情報を生成する。第 2 統合情報生成部 3 2 2 は、第 1 統合情報に含まれる各情報と、検索情報に含まれる各情報とを含む第 2 統合情報を生成する。第 2 統合情報生成部 3 2 2 は、各特許文献にファミリー出願が存在する場合、ファミリー出願に係る各特許文献から生成した第 2 統合情報を一つの第 2 テーブル内で相互に対応付けて記憶する。第 2 統合情報生成部 3 2 2 は、一つの第 2 テーブル内で、ファミリー出願に係る特許文献毎に第 2 統合情報のレコードを設定する。

## 【 0 1 3 1 】

このように、第 2 統合情報生成部 3 2 2 は、照会サーバ 1 0 0 が形式情報を取得した各特許文献に係る第 1 統合情報が完成した後に、各特許文献について、第 2 統合情報を生成する。即ち、第 2 統合情報生成部 3 2 2 は、照会サーバ 1 0 0 が形式情報を取得した各特許文献に係る第 1 統合情報の生成が完了するまで第 2 統合情報の生成を開始しない。これにより、第 2 統合情報生成部 3 2 2 は、第 2 統合情報を効率良く生成することができる。

## 【 0 1 3 2 】

次に、第 2 統合情報送信部 3 2 3 は、第 2 統合情報を、第 3 通信装置 3 0 1 を介して照会サーバ 1 0 0 に送信する（ステップ S 1 1 6）。

## 【 0 1 3 3 】

照会サーバ 1 0 0 の第 3 統合情報生成部 1 2 2 は、第 1 通信装置 1 0 1 を介して管理サーバ 3 0 0 から第 2 統合情報を受信した場合、受信した第 2 統合情報と、各特許文献に含まれる内容情報のテキストデータとを統合した第 3 統合情報を生成する（ステップ S 1 1 7）。第 3 統合情報生成部 1 2 2 は、照会サーバ 1 0 0 が形式情報を取得した各特許文献について、ステップ S 1 0 1 で受信した内容情報からテキストデータを抽出する。テキストデータは、内容情報に含まれるデータの一例である。なお、第 3 統合情報生成部 1 2 2 は、抽出したテキストデータを管理システム 1 内の共通形式のフォーマットに変換する。

## 【 0 1 3 4 】

照会サーバ 1 0 0 は、各特許文献について、第 2 統合情報及び内容情報の各データ項目が所定の順序で並べられた第 3 テーブルを第 1 記憶装置 1 1 0 に記憶する。第 3 統合情報生成部 1 2 2 は、各特許文献について、文献番号をキーとして、管理サーバ 3 0 0 から受信した第 2 統合情報に含まれる各データと、内容情報から抽出したテキストデータとを、第 3 テーブルの対応する位置に格納する。これにより、第 3 統合情報生成部 1 2 2 は、照会サーバ 1 0 0 が形式情報を取得した各特許文献について、第 3 統合情報を生成する。

## 【 0 1 3 5 】

図 2 0 は、第 3 統合情報（第 3 テーブル）のデータ構造の一例を示す模式図である。

## 【 0 1 3 6 】

図 2 0 に示すように、第 3 統合情報には、第 2 統合情報に含まれる各情報に加えて、テキストデータ、二次データ及び管理情報等が含まれる。図 2 0 に示すように、第 3 統合情報には、上記した項目毎に、データ型及び一又は複数のレコードが含まれる。このように、第 2 テーブルには、第 2 統合情報及び内容情報の各データ項目が所定の順序で並べられている。また、第 3 統合情報のデータ構造は、図 1 0 ~ 図 1 3 に示した管理テーブル 5 1 1 のデータ構造と同様である。

10

20

30

40

50



## 【 0 1 3 7 】

第3統合情報生成部122は、受信した各第2統合情報と、抽出したテキストデータの  
内、文献番号が同一である第2統合情報及びテキストデータの組合せを統合（マージ）す  
る。第3統合情報生成部122は、各特許文献にファミリー出願が存在する場合、ファミ  
リ出願に係る各特許文献から生成した第3統合情報を一つの第3テーブル内で相互に対  
応付けて記憶する。第3統合情報生成部122は、一つの第3テーブル内で、ファミリー  
出願に係る特許文献毎に第3統合情報のレコードを設定する。また、第3統合情報生成部  
122は、各特許文献に係るファミリー出願にファミリーIDを割り当てるとともに、ファミ  
リ出願に対応する特許文献の中で特定の特許文献の代表文献フラグを有効に設定し  
、ファミリーID及び代表文献フラグを二次データとして記憶する。また、第3統合情報  
生成部122は、その特許文献の最新の更新日を更新日に設定する。また、第3統合情報  
生成部122は、自サーバ内で各特許文献のテキストファイルが格納されたアドレスを格  
納アドレスに設定する。さらに、第3統合情報生成部122は、第3統合情報の送信先の  
検索サーバ500の識別情報を検索サーバ名に設定し、自サーバの識別情報を照会サー  
バ名に設定する。そして、第3統合情報生成部122は、更新日、格納アドレス、検索サー  
バ名及び照会サーバ名を管理情報として記憶する。

## 【 0 1 3 8 】

このように、第3統合情報生成部122は、照会サーバ100が形式情報を取得した各  
特許文献に係る第2統合情報が完成した後に、各特許文献について、第3統合情報を生成  
する。即ち、第3統合情報生成部122は、照会サーバ100が形式情報を取得した各特  
許文献に係る第2統合情報の生成が完了するまで第3統合情報の生成を開始しない。これ  
により、第3統合情報生成部122は、第3統合情報を効率良く生成することができる。

## 【 0 1 3 9 】

次に、第3統合情報送信部123は、第3統合情報を、検索用データベース600に一  
括して登録するように、第1通信装置101を介して検索サーバ500に送信する（ステ  
ップS118）。

## 【 0 1 4 0 】

検索サーバ500の第3統合情報記憶制御部521は、第5通信装置501を介して照  
会サーバ100から第3統合情報を受信した場合、照会サーバ100から送信された第3  
統合情報を一括して管理テーブル511に記憶する（ステップS119）。これにより、  
第3統合情報記憶制御部521は、第3統合情報を検索用データベース600に一括して  
登録する。第3統合情報記憶制御部521は、第3統合情報に含まれる各項目を管理テー  
ブル511の対応する項目の位置に記憶する。上記したように、第3統合情報のデータ構  
造は、管理テーブル511のデータ構造と同様である。そのため、第3統合情報記憶制  
御部521は、第3統合情報を加工することなく、単純に管理テーブル511に追加する  
ことにより、管理テーブル511を簡易に更新できる。したがって、第3統合情報記憶制  
御部521は、管理テーブル511の更新処理の処理負荷を軽減させ、且つ、処理時間を低  
減させることが可能となる。

## 【 0 1 4 1 】

また、図10～図13に示すように、第3統合情報記憶制御部521は、第3統合情報  
の各レコード（データ項目）に対してインデックスを設定する。第3統合情報記憶制御部  
521は、各特許文献にファミリー出願が存在する場合、ファミリー出願に係る特許文献  
から生成された第3統合情報に含まれる各情報のデータ項目のインデックスとして、ファミ  
リ出願に係る特許文献で共通のインデックスを設定する。例えば、第3統合情報記憶  
制御部521は、代表文献フラグが有効に設定された特許文献について設定したインデッ  
クスをファミリー出願に係る特許文献で共通のインデックスとして使用する。これにより  
、検索サーバ500は、特許文献の検索を行う際に、インデックスを用いて短時間に検索  
を行うことができる。以上により、更新処理に係る動作シーケンスは終了する。

## 【 0 1 4 2 】

このように、照会サーバ100は、第2統合情報を含む第3統合情報を検索用データベ

ース600に登録する。なお、照会サーバ100の代わりに、管理サーバ300が、第3統合情報を生成し、検索用データベース600に登録してもよい。または、管理サーバ300が、第2統合情報を検索用データベース600に登録し、照会サーバ100又は管理サーバ300が、内容情報を検索用データベース600に登録してもよい。

【0143】

また、上記したように、検索情報生成処理は、形式情報の取得及び書誌情報の取得より長時間を有する。そのため、検索情報生成処理は、形式情報の取得及び書誌情報の取得の開始時である第1タイミングより前に開始されてもよい。その場合、検索情報生成処理は、第1統合情報が完成する前又は第1統合情報の生成が開始される前に完了してもよい。

【0144】

また、管理サーバ300は、第1統合情報が完成した時点で検索情報生成処理が完了していない場合、即時にAIサーバ400に検索要求情報を送信し、第1統合情報が完成する前に生成された検索情報のみをAIサーバ400から取得してもよい。その場合、管理サーバ300は、第1統合情報が完成した後に、第1統合情報に含まれる各データと、第1統合情報が完成する前に生成された検索情報に含まれる各データとを、第2テーブルの対応する位置に格納する。これにより、管理サーバ300は、第1統合情報と、第1統合情報が完成する前に生成された検索情報とを統合した第2統合情報を生成する。なお、管理サーバ300は、第1統合情報が完成した時点で生成されていなかった検索情報については、次に更新処理が実行される期間に第2統合情報を生成する。

【0145】

図21は、AIサーバ400における検索情報生成処理の動作の例を示すフローチャートである。

【0146】

図21に示す検索情報生成処理は、図14に示す更新処理のステップS106で実行される。以下のステップS201～S206の処理は、特許文献毎に実行される。

【0147】

まず、検索情報生成部421は、AI技術を用いて、各特許文献の書誌情報及び内容情報を翻訳した翻訳文を検索情報として生成する(ステップS201)。検索情報生成部421は、各特許文献、即ち検索情報を生成する対象の特許文献の内容情報に示される発明の内容が、対象特許庁における使用言語と異なる言語で記載されているか否かを判定する。発明の内容が使用言語と異なる言語で記載されている場合、検索情報生成部421は、その言語で記載された発明の内容を使用言語に翻訳するための機械翻訳エンジンである第1学習モデル411を利用して、その発明の内容の使用言語による翻訳文を生成する。翻訳文は、翻訳データの一例である。

【0148】

使用言語、即ち翻訳後の言語は、日本語、英語、ドイツ語、フランス語、中国語、韓国語等の任意の言語であり、特許文献が記載されている言語と異なる言語であればどのような言語でもよい。このような機械翻訳エンジンとして任意の翻訳エンジンを利用してよい。また、AIサーバ400は、公知のAI技術を利用して、様々な言語で記載された学習用特許文献を用いて事前学習することにより、第1学習モデル411を独自に生成してもよい。なお、検索情報生成部421は、各特許文献の書誌情報及び内容情報を翻訳した翻訳文の作成を要求する作成要求信号を他のサーバに送信し、翻訳文を他のサーバから受信することにより取得してもよい。また、検索情報生成部421は、外部の翻訳業者により作成された翻訳文をUSB(Universal Serial Bus)等のインタフェース規格に従った不図示のインタフェース装置から入力することにより取得してもよい。

【0149】

次に、検索情報生成部421は、各特許文献の内容情報に基づいて、その特許文献の技術分野の大分類を特定し、特定した大分類を分類情報として生成する(ステップS202)。なお、検索情報生成部421は、各特許文献の内容情報及び書誌情報に基づいて、その特許文献の技術分野の大分類を特定してもよい。検索情報生成部421は、各特許文献

10

20

30

40

50

において、対象特許庁が割り当てべき技術分野の小分類が規定されているか否かを判定する。その小分類が規定されていない場合、検索情報生成部 4 2 1 は、まず、以下の四つの方法により、技術分野の大分類を特定する。

【 0 1 5 0 】

検索情報生成部 4 2 1 は、第 1 の方法として、各特許文献に含まれる対象特許庁以外の特許庁により規定された技術分野の分類に基づいて、対象特許庁により規定される技術分野の大分類を特定する。各特許文献を発行した特許庁が米国特許商標庁又は欧州特許庁等である場合、各特許文献に含まれる技術分野の分類として、IPC、CPC等が使用可能である。また、各特許文献を発行した特許庁が日本国特許庁である場合、各特許文献に規定された技術分野の分類として、日本国特許庁が割り当ててるテーマコード、FI及びFターム等が使用可能である。 10

【 0 1 5 1 】

AIサーバ 4 0 0 は、各国特許庁により規定される技術分野の分類毎に、対象特許庁により規定される技術分野の大分類を対応付けて第 5 記憶装置 5 1 0 に予め設定しておく。検索情報生成部 4 2 1 は、各特許文献から、何れかの国の特許庁により規定された技術分野の分類を特定し、特定した技術分野の分類に対応付けて設定された対象特許庁により規定される技術分野の大分類を特定する。

【 0 1 5 2 】

検索情報生成部 4 2 1 は、第 2 の方法として、ファミリー出願間の技術分野の分類の対応関係の統計データに基づいて、技術分野の大分類を特定する。AIサーバ 4 0 0 は、過去に出願されたファミリー出願に係る特許文献の組合せに基づいて統計データを事前に生成しておく。AIサーバ 4 0 0 は、対象特許庁以外の各国特許庁により規定される技術分野の分類毎に、その分類が割り当てられた特許文献のファミリー出願の内、対象特許庁へのファミリー出願に係る特許文献を抽出する。そして、AIサーバ 4 0 0 は、対象特許庁以外の各国特許庁により規定される技術分野の分類毎に、抽出した各特許文献において、対象特許庁により規定される各分類が割り当てられている数又は比率を統計データとして算出して記憶しておく。検索情報生成部 4 2 1 は、各特許文献が、対象特許庁以外の各国特許庁から収集した特許文献である場合、その特許文献から、その国の特許庁により規定された技術分野の分類を特定する。そして、検索情報生成部 4 2 1 は、特定した技術分野の分類に対応付けて記憶された統計データが閾値以上である分類の大分類を、取得した特許文献の技術分野の大分類として特定する。 20 30

【 0 1 5 3 】

例えば、対象特許庁が日本国特許庁である場合、AIサーバ 4 0 0 は、特定のIPC又はCPC等が割り当てられた出願のファミリー出願において特定のFI又はFターム等が割り当てられた数又は比率を統計データとして算出しておく。検索情報生成部 4 2 1 は、各特許文献において割り当てられたIPC又はCPCを特定し、特定したIPC又はCPCに対応付けられた統計データが閾値以上であるFI又はFタームのテーマコードを、各特許文献の技術分野の大分類として特定する。

【 0 1 5 4 】

検索情報生成部 4 2 1 は、第 3 の方法として、概念検索を利用して、技術分野の大分類を特定する。検索情報生成部 4 2 1 は、後述する概念検索を利用して、取得した特許文献と類似する、対象特許庁に出願された特許文献を所定数抽出する。概念検索では、例えば各特許文献の内容情報に含まれる各用語の出現頻度が近似する特許文献が抽出される。検索情報生成部 4 2 1 は、抽出した各特許文献において、割り当てられている数又は比率が閾値以上である技術分野の分類の大分類を、各特許文献の技術分野の大分類として特定する。 40

【 0 1 5 5 】

また、検索情報生成部 4 2 1 は、第 4 の方法として、内容情報に基づいて、機械学習(SVM等)によって大分類を特定する。

【 0 1 5 6 】

検索情報生成部 4 2 1 は、第 1 ~ 第 4 の方法で特定した全ての大分類を、各特許文献の技術分野の大分類として特定する。なお、検索情報生成部 4 2 1 は、第 1 ~ 第 4 の方法の内の一つ又は二つの方法のみにより、各特許文献の技術分野の大分類を特定してもよい。

【 0 1 5 7 】

次に、検索情報生成部 4 2 1 は、A I 技術を用いて、各特許文献の技術分野の小分類を特定し、特定した小分類を分類情報として生成する（ステップ S 2 0 3 ）。

【 0 1 5 8 】

検索情報生成部 4 2 1 は、ステップ S 2 0 2 で特定された大分類に対応する第 2 学習モデル 4 1 2 を用いて、各特許文献の技術分野の小分類を特定する。各第 2 学習モデル 4 1 2 は、対象特許庁における使用言語で記載された複数の学習用特許文献を用いて、各特許文献の内容情報に関する情報が入力された場合に、その特許文献の技術分野に関する情報を出力するように事前学習されている。特に、各第 2 学習モデル 4 1 2 は、各特許文献の内容情報から算出された特徴量が入力された場合に、その特許文献の技術分野の小分類に関する情報を出力するように事前学習されている。

【 0 1 5 9 】

例えば、各第 2 学習モデル 4 1 2 は、小分類毎に生成され、S V M ( Support Vector Machine ) を用いて学習される。各第 2 学習モデル 4 1 2 は、各特許文献の内容情報から算出された特徴量が入力されたときに、その特許文献が、対応する小分類に合致している場合は + 1 を出力し、合致していない場合は - 1 を出力するように学習される。即ち、各第 2 学習モデル 4 1 2 は、対応する小分類に合致している特許文献から算出された特徴量と、対応する小分類に合致していない特許文献から算出された特徴量とを識別する識別平面を含む。各第 2 学習モデル 4 1 2 は、入力された特徴量が、その識別平面に対して、対応する小分類に合致している側に位置する場合は + 1 を出力し、対応する小分類に合致している側に位置する場合は - 1 を出力する。なお、各第 2 学習モデル 4 1 2 は、入力された特徴量が、その識別平面に対して、対応する小分類に合致している側に位置する場合、その特徴量の識別平面からの距離（マージン）を正規化したスコア値を出力してもよい。パラメータ調整及び閾値調整は、公知の調整方法により実行される。

【 0 1 6 0 】

特徴量として、例えば T F - I D F ( Term Frequency Inverse Document Frequency ) が使用される。A I サーバ 4 0 0 は、形態素解析技術を用いて、各学習用特許文献内の内容情報を単語（形態素）毎に分解し、各単語の出現頻度と逆文書頻度とを算出して、T F - I D F を算出する。なお、逆文書頻度は、同一のテーマコードを有する特許文献から算出される。また、各単語の出現頻度の次元数を削減するために、下限閾値以下又は上限閾値以上である形態素は除去されてもよい。また、A I サーバ 4 0 0 は、墨付き括弧で囲まれた「技術分野」又は「背景技術」等の特許文献における特定のフォーマットに対応する各用語を除去してから文書を分解してもよい。

【 0 1 6 1 】

なお、特徴量として、B a g o f W o r d s 等の T F - I D F 以外の特徴量が用いられてもよい。また、特徴量として、各形態素（単語）、文、段落もしくは文献の分散表現、テキスト及び / 又は I P C 等の技術分野の分類情報が用いられてもよい。また、特徴量として、各特許文献の内容情報に含まれる図面内の画像から抽出された対象物の角又は交点等の特徴点の数又は分布等に関する情報が用いられてもよい。

【 0 1 6 2 】

また、第 2 学習モデル 4 1 2 は、ロジスティック回帰、M L P ( Multilayer Perceptron )、R N N ( Recurrent Neural Network )、C N N ( Convolutional Neural Network )、N A M ( Neural Attention Model ) 等の他の公知の機械学習技術を用いて学習されてもよい。また、第 2 学習モデル 4 1 2 は、アンサンブル学習等の手法を用い、複数の機械学習技術を組み合わせて学習されてもよい。その場合には、複数の機械学習を組み合わせるパラメータ自体も機械学習で求めてもよい。

【 0 1 6 3 】

10

20

30

40

50

検索情報生成部 4 2 1 は、A I サーバ 4 0 0 による事前学習処理と同様にして、各特許文献から特徴量を算出する。特徴量が T F - I D F である場合、検索情報生成部 4 2 1 は、形態素解析技術を用いて、特許文献内の内容情報を単語毎に分解し、各単語の出現頻度と逆文書頻度とを算出して、T F - I D F を算出する。検索情報生成部 4 2 1 は、算出した特徴量を、ステップ S 2 0 3 で特定された大分類に属する各小分類に対応する各第 2 学習モデル 4 1 2 に入力し、各第 2 学習モデル 4 1 2 からの出力値を取得する。検索情報生成部 4 2 1 は、出力値が閾値以上である第 2 学習モデル 4 1 2 に対応する小分類を、取得した特許文献の技術分野の小分類として特定し、特定した技術分野の小分類を検索情報として生成する。

【 0 1 6 4 】

検索情報を生成する対象の特許文献が対象特許庁における使用言語と異なる言語で記載されている場合、検索情報生成部 4 2 1 は、その特許文献の使用言語による翻訳文から特徴量を算出する。この特徴量は、特許文献の使用言語による翻訳文に関する情報の一例である。この場合、検索情報生成部 4 2 1 は、特許文献の使用言語による翻訳文から算出された特徴量を第 2 学習モデル 4 1 2 に入力することにより、その特許文献の技術分野に関する情報を取得する。これにより、管理システム 1 は、様々な言語の特許文献に対して、各特許文献が記載された言語の種別に関わらず、技術分野を適切に付与することができる。

【 0 1 6 5 】

なお、検索情報生成部 4 2 1 は、対象特許庁における使用言語で記載され且つ技術分野の小分類が付与されていない特許文献についても、各第 2 学習モデル 4 1 2 に入力して、その特許文献の小分類を特定してもよい。

【 0 1 6 6 】

また、検索情報生成部 4 2 1 は、分類付与の根拠となる単語（形態素）、文又は段落等を特定してもよい。例えば、第 2 学習モデル 4 1 2 が、T F - I D F 等を特徴量として、S V M 又はロジスティック回帰等を用いて学習される場合、その学習時に特徴量内の各要素の重みが定められる。検索情報生成部 4 2 1 は、特許文献から算出された特徴量内の各要素に、各要素に定められた重みを乗算した乗算値を算出し、乗算値が所定閾値以上である要素に対応する単語を、分類付与の根拠となる単語として特定する。なお、検索情報生成部 4 2 1 は、特許文献内の文又は段落毎に、各文又は段落に含まれる単語に係る乗算値の総和値を算出し、総和値が所定閾値以上である要素に対応する文又は段落を、分類付与の根拠となる文又は段落として特定してもよい。検索情報生成部 4 2 1 は、特定した単語、文又は段落を示す情報を、各特許文献及び技術分野を示す情報と関連付けて照会サーバ 1 0 0 に通知する。照会サーバ 1 0 0 は、端末装置 1 0 から照会要求信号を受信した場合に、指定された文献番号に対応する特許文献のテキストデータ及びイメージデータとともに、検索情報生成部 4 2 1 により特定された技術分野、単語、文又は段落を端末装置 1 0 に送信する。これにより、端末装置 1 0 の利用者は、技術分野が特定された根拠となる単語、文又は段落から、特定された技術分野の妥当性を検討することができ、必要に応じて技術分野を修正することができる。

【 0 1 6 7 】

また、検索情報生成部 4 2 1 は、特許文献内のブロック毎に、技術分野の小分類に関する情報を算出し、段落毎に算出した情報に基づいて、技術分野の小分類を特定してもよい。ブロックは、文又は段落等である。その場合、例えば、第 2 学習モデル 4 1 2 は、小分類毎に生成され、学習用特許文献に含まれる各ブロックから算出された B a g o f W o r d s 等を特徴量として、S V M を用いて学習される。各第 2 学習モデル 4 1 2 は、入力された特徴量が、識別平面に対して、対応する小分類に合致している側に位置する場合に、その特徴量の識別平面からの距離（マージン）を正規化したスコア値を出力するように学習される。なお、特定の小分類に対応する学習用特許文献に含まれるブロックの内、その小分類と関連しないブロックが、特定の小分類に対応しない学習用サンプルとして使用されてもよい。

10

20

30

40

50

## 【 0 1 6 8 】

検索情報生成部 4 2 1 は、特許文献に含まれる各ブロックから算出した特徴量を、対応する第 2 学習モデル 4 1 2 に入力し、各第 2 学習モデル 4 1 2 から出力されたスコア値を取得する。検索情報生成部 4 2 1 は、スコア値が閾値以上であるブロックが所定数（例えば 1 つ）以上存在する場合、その第 2 学習モデル 4 1 2 に対応する小分類を、その特許文献の技術分野の小分類として特定する。検索情報生成部 4 2 1 は、その特許文献に係る出力値の最大値に基づいて第 2 閾値を設定し、スコア値が第 2 閾値以上であるブロックが所定数以上存在する場合、その第 2 学習モデル 4 1 2 に対応する小分類を、その特許文献の技術分野の小分類として特定してもよい。第 2 閾値は、例えば最大値の所定倍（例えば 0 . 5 倍）に設定される。これらにより、検索情報生成部 4 2 1 は、より精度良く、特許文献の技術分野を特定することができる。

## 【 0 1 6 9 】

この場合の特徴量として、TF - IDF、各形態素（単語）、又は、文、段落もしくは文献の分散表現等が用いられてもよい。また、第 2 学習モデル 4 1 2 は、ロジスティック回帰、MLP、RNN、CNN、NAM等の他の公知の機械学習技術を用いて、又は、複数の機械学習技術を組み合わせて学習されてもよい。

## 【 0 1 7 0 】

また、この場合、検索情報生成部 4 2 1 は、特許文献に含まれるブロック毎に、閾値を異ならせてもよい。例えば、検索情報生成部 4 2 1 は、特許請求の範囲又は発明の概要等に含まれるブロックに対応する閾値を、他のブロックに対応する閾値より小さくする。これにより、検索情報生成部 4 2 1 は、特許請求の範囲又は発明の概要に含まれる用語を、より優先的に考慮して、特許文献の技術分野を特定することができる。

## 【 0 1 7 1 】

また、検索情報生成部 4 2 1 は、特許文献の内容情報全体から算出されたスコア値と、特許文献内の各ブロックから算出されたスコア値とに基づいて、技術分野の小分類を特定してもよい。その場合、検索情報生成部 4 2 1 は、特許文献の内容情報全体から算出された特徴量を、学習用特許文献の内容情報全体を用いて学習された学習モデルに入力して第 1 スコア値を取得する。また、検索情報生成部 4 2 1 は、特許文献内の各ブロックから算出された特徴量を、学習用特許文献内の各ブロックを用いて学習された学習モデルに入力して第 2 スコア値を取得する。また、AIサーバ 4 0 0 は、第 1 スコア値及び第 2 スコア値が入力された場合に、その特許文献が、対応する小分類に合致するか否かを出力するように事前学習された学習モデルを第 4 記憶装置 4 1 0 に事前に記憶しておく。検索情報生成部 4 2 1 は、取得した第 1 スコア値及び第 2 スコア値をその学習モデルに入力することにより、その特許文献が、対応する小分類に合致するか否かを判定する。なお、検索情報生成部 4 2 1 は、第 1 スコア値及び第 2 スコア値の和又は重み付き和が所定値以上であるか否かにより、その特許文献が、対応する小分類に合致するか否かを判定してもよい。これにより、検索情報生成部 4 2 1 は、より精度良く、特許文献の技術分野を特定することができる。

## 【 0 1 7 2 】

また、第 2 学習モデル 4 1 2 は、各特許文献から算出された特徴量が入力された場合に、その特許文献が最も合致する小分類を示す情報を出力するように事前学習されてもよい。その場合、第 2 学習モデル 4 1 2 は、複数の学習用特許文献を用いて、例えばディープラーニングにより学習される。ニューラルネットワークである学習モデルは、入力層、中間層及び出力層から構成される多層構造を有する。入力層の各ノードには、特徴量として、各特許文献から算出されたTF - IDF等の各要素等が入力される。中間層の各ノードには、入力層の各ノードから出力された値の、所定の重みに基づく重み付け和が入力される。出力層の各ノードには、中間層の各ノードから出力された値の、所定の重みに基づく重み付け和が入力される。出力層は、例えば、入力された値をそのまま出力する。各重みは、出力層によって出力された値と学習用特許文献に割り当てられた技術分野の小分類を示す値との差分が小さくなるように設定される。重みの設定には、例えば、誤差逆伝搬法

などの公知の手法が使用される。検索情報生成部 4 2 1 は、算出した特徴量を各第 2 学習モデル 4 1 2 に入力し、各第 2 学習モデル 4 1 2 からの出力値に対応する小分類を、取得した特許文献の技術分野の小分類として特定する。

【 0 1 7 3 】

このように、検索情報生成部 4 2 1 は、学習モデルを用いて、各特許文献の書誌情報又は内容情報に基づいて、各特許文献の技術分野又は翻訳データ等を、各特許文献から推定される検索情報として生成する。各データベースには、人手によって技術分野が割り当てられた特許文献が多数存在しており、検索情報生成部 4 2 1 は、多数の教師データを利用して、各特許文献の技術分野を高精度に特定することができる。

【 0 1 7 4 】

特に、検索情報生成部 4 2 1 は、発明の内容が対象特許庁における使用言語で記載されていない特許文献について、発明の内容がその使用言語に翻訳された翻訳文を生成する。検索情報生成部 4 2 1 は、発明の内容がその使用言語で記載された特許文献についてはその発明の内容に基づき、発明の内容がその使用言語で記載されていない特許文献については発明の内容がその使用言語に翻訳された翻訳文に基づいて、学習モデルを用いて、各特許文献の分類情報を生成する。即ち、検索情報生成部 4 2 1 は、外国語の特許文献を、学習モデルが対応する言語に翻訳してから、その技術分野の小分類を特定する。これにより、AIサーバ 4 0 0 は、複数の言語毎に学習モデルを用意する必要がなくなり、事前学習に要する担当者の作業量を削減するとともに、第 5 記憶装置 5 1 0 の記憶容量の削減を図ることができる。また、AIサーバ 4 0 0 は、特定の言語に絞って学習モデルを生成すればよく、十分な時間及び労力をかけて学習モデルを生成することができ、結果として、学習モデルの精度を向上させることが可能となる。特に、近年では、ファミリー出願に係る特許文献対を用いて学習することにより、AI技術を用いた特許文献の翻訳技術が向上しており、検索情報生成部 4 2 1 は、高品質な翻訳文を取得できる。そのため、検索情報生成部 4 2 1 は、高品質な翻訳文を利用して、技術分野の小分類を高精度に特定することができる。

【 0 1 7 5 】

なお、AIサーバ 4 0 0 は、対象特許庁における使用言語以外の複数の言語毎に、各言語で記載された学習用特許文献を用いて事前学習した学習モデルを第 4 記憶装置 4 1 0 に記憶しておいてもよい。その場合、AIサーバ 4 0 0 は、使用言語以外の言語で記載された学習用特許文献として、使用言語で記載された特許文献から翻訳した翻訳文を用いてもよい。また、学習モデルは、その学習用特許文献のファミリー出願に割り当てられた技術分野の小分類を用いて学習されてもよい。この場合、検索情報生成部 4 2 1 は、取得した特許文献を翻訳することなく、その特許文献が記載された言語に対応した学習モデルを用いて、技術分野の小分類を特定する。これにより、検索情報生成部 4 2 1 は、短時間で技術分野の小分類を特定ことができ、検索情報生成処理に係る処理時間を短縮させるとともに、処理負荷を軽減させることができる。

【 0 1 7 6 】

また、検索情報生成部 4 2 1 は、各特許文献の技術分野の大分類を特定してから、その大分類に対応する学習モデルを用いて、技術分野の小分類を特定する。これにより、検索情報生成部 4 2 1 は、各特許文献を、全ての小分類に対応する学習モデルに適用させる必要がなくなり、検索情報生成処理に係る処理時間を短縮させるとともに、処理負荷を軽減させることができる。なお、検索情報生成部 4 2 1 は、各特許文献の技術分野の大分類を特定することなく、各特許文献の内容情報から技術分野の小分類を特定してもよい。

【 0 1 7 7 】

次に、検索情報生成部 4 2 1 は、各特許文献の内容情報から第 2 特徴ベクトルを生成する(ステップ S 2 0 4)。第 2 特徴ベクトルとして、例えば内容情報から算出された TF-IDF が使用される。なお、第 2 特徴ベクトルとして、Bag of Words 又は BM 2 5 等の TF-IDF 以外の特徴ベクトルが用いられてもよい。また、第 2 特徴ベクトルとして、各形態素(単語)、文、段落もしくは文献の分散表現が用いられてもよい。

分散表現として、例えば `Word2Vec`、`Doc2Vec`、`SDCV` (Sparse Composite Document Vectors) 等が用いられる。

【0178】

次に、検索情報生成部421は、各特許文献について、第2特徴ベクトルから、それぞれ次元数が異なる複数の第1特徴ベクトルを生成する(ステップS205)。検索情報生成部421は、それぞれ異なる複数のハッシュ関数LSH(Locality Sensitive Hashing)を用いて各特許文献の第2特徴ベクトルを変換したハッシュ値を各要素とする特徴ベクトルを第1特徴ベクトルとして生成する。

【0179】

ハッシュ関数LSHは、特徴ベクトルが近似している程、ハッシュ値が近似するように設定される関数である。各ハッシュ関数は、以下の式により定義される。

【数1】

$$h(v) = \left\lfloor \frac{a \cdot v + b}{w} \right\rfloor$$

ここで、 $v$ は、第2特徴ベクトルであり、 $a$ は、ガウス分布を例とする $p$ 安定分布から選択された、第2特徴ベクトル $v$ の次元数 $m$ と同数の確率変数からなる $m$ 次元ベクトルであり、 $b$ は平空間 $[0, W]$  ( $W > 0$ ) から一様ランダムに選ばれた実数である。第2特徴ベクトル $v$ の特徴量空間は、 $h(v)$ によって、ベクトル $a$ と直交する等間隔の超平面で分割される。

【0180】

第1特徴ベクトル $g_k(v)$ は、以下の式により定義される。

$$g_k(v) = (h_1(v), h_2(v), \dots, h_k(v))$$

ここで、 $h_1(v) \sim h_k(v)$ は、ランダムに設定された、それぞれ異なる $a$ 、 $b$ により定義されるハッシュ関数 $h(v)$ のハッシュ値である。即ち、第1特徴ベクトル $g_k(v)$ は、 $k$ 次元の特徴ベクトルであり、 $k$ 個の等間隔な超平面集合で第2特徴ベクトル $v$ の特徴量空間を分割した各部分空間(バケツ)の内、その特許文献に対応する部分空間(バケツ)を表している。

【0181】

検索情報生成部421は、次元数 $k$ がそれぞれ異なる二以上の所定数の第1特徴ベクトル $g_k(v)$ の式を設定する。検索情報生成部421は、例えば次元数 $k$ が1、2、4、8、16、32、64、128、256である9個の第1特徴ベクトル $g_1(v)$ 、 $g_2(v)$ 、 $g_4(v)$ 、 $g_8(v)$ 、 $g_{16}(v)$ 、 $g_{32}(v)$ 、 $g_{64}(v)$ 、 $g_{128}(v)$ 、 $g_{256}(v)$ を設定する。

【0182】

検索情報生成部421は、設定した所定数の第1特徴ベクトル $g_k(v)$ の各式に第2特徴ベクトル $v$ を代入することにより、それぞれ次元数が異なる複数の第1特徴ベクトルを生成する。第1特徴ベクトル $g_k(v)$ の次元数 $k$ は、第2特徴ベクトルの特徴量空間を各部分空間に分割する分割数に対応し、次元数 $k$ が大きいく程、分割数は大きくなり、各第1特徴ベクトル $g_k(v)$ が表す情報は大きくなる。

【0183】

次に、検索情報生成部421は、情報処理装置15が所定期間に収集した全ての特許文献について処理が完了したか否かを判定する(ステップS206)。まだ処理が完了していない特許文献が存在する場合、検索情報生成部421は、処理をステップS201に戻し、ステップS201~S206の処理を繰り返す。一方、全ての特許文献について処理が完了した場合、検索情報生成部421は、一連のステップを終了する。

【0184】

なお、AIサーバ400は、さらに、各特許文献の内容情報のテキスト又は図面が入力



された場合に、その特許文献の特徴ベクトル、キーワード又は図面のメタデータを出力するように事前学習された学習モデルを有してもよい。その場合、学習モデルは、例えばディープラーニングを用いて学習され、各重みは、出力層によって出力された値と学習用特許文献に割り当てられた特徴ベクトル、キーワード又はメタデータを示す値との差分が小さくなるように設定される。検索情報生成部 4 2 1 は、各特許文献の内容情報のテキスト又は図面を各学習モデルに入力し、各学習モデルからの出力値に対応する特徴ベクトル、キーワード又はメタデータを、各特許文献の特徴ベクトル、キーワード又はメタデータに決定する。検索情報生成部 4 2 1 は、決定した特徴ベクトル、キーワード又はメタデータを検索情報として生成する。

**【 0 1 8 5 】**

また、検索情報生成部 4 2 1 は、各特許文献にファミリー出願が存在する場合、ファミリー出願に係る各特許文献について共通の検索情報を生成してもよい。その場合、検索情報生成部 4 2 1 は、代表文献フラグが有効に設定された特許文献について生成した検索情報を、ファミリー出願に係る特許文献で共通の検索情報として使用する。即ち、検索情報生成部 4 2 1 は、特定の出願に係る特許文献について生成した分類情報を含む検索情報を、その特定の出願のファミリー出願に係る特許文献の検索情報として使用する。これにより、検索情報生成部 4 2 1 は、より短時間で技術分野の小分類を設定することができ、検索情報生成処理に係る処理時間を短縮させるとともに、処理負荷を軽減させることができる。

**【 0 1 8 6 】**

図 2 2 は、図 1 4 に示した更新処理内の各処理の実行タイミングについて説明するための模式図である。

**【 0 1 8 7 】**

図 2 2 に示す例では、更新処理は、7 日間（1 週間）周期で実行される。形式情報の取得処理、書誌情報の取得処理及び検索情報の生成処理が 1 日目に同時に開始され、形式情報の取得処理及び書誌情報の取得処理は 3 日目に完了するが、検索情報の生成処理は 4 日目まで完了しない。形式情報及び書誌情報を用いた第 1 統合情報の生成処理が 3 日目に開始され、4 日目に完了する。そして、第 1 統合情報及び検索情報を用いた第 2 統合情報の生成処理が 5 日目に開始され、6 日目に完了する。最後に、第 2 統合情報及び内容情報を用いた第 3 統合情報の生成処理が 6 日目に開始され、7 日目に完了する。

**【 0 1 8 8 】**

このように、管理サーバ 3 0 0 は、生成に多大な時間を要する検索情報の完成を待つことなく、第 1 統合情報の生成処理を開始させる。これにより、管理サーバ 3 0 0 は、検索情報が完成するまでに第 1 統合情報を生成しておき、生成した第 1 統合情報及び検索情報を用いて、効率良く短時間に第 2 統合情報を生成できる。

**【 0 1 8 9 】**

また、管理サーバ 3 0 0 は、各特許文献の形式情報及び書誌情報を取得するたびに第 1 統合情報を生成するのではなく、所定期間に情報処理装置 1 5 が収集した複数の特許文献の形式情報及び書誌情報の取得が完了してから第 1 統合情報を生成する。これにより、各サーバにおける形式情報及び書誌情報の送受信処理を含む、第 1 統合情報の生成処理に係るオーバーヘッドが低減される。同様に、管理サーバ 3 0 0 は、所定期間に情報処理装置 1 5 が収集した複数の特許文献の第 1 統合情報及び検索情報が完成してから第 2 統合情報を生成して照会サーバ 1 0 0 に送信する。これにより、各サーバにおける検索情報及び第 2 統合情報の送受信処理を含む、第 2 統合情報の生成処理に係るオーバーヘッドが低減される。さらに、照会サーバ 1 0 0 は、所定期間に情報処理装置 1 5 が収集した複数の特許文献の第 2 統合情報が完成してから第 3 統合情報を生成して検索サーバ 5 0 0 に送信する。これにより、各サーバにおける第 3 統合情報の送受信処理を含む、照会サーバ 1 0 0 による第 3 統合情報の生成処理及び検索サーバ 5 0 0 における管理テーブルの更新処理に係るオーバーヘッドが低減される。

**【 0 1 9 0 】**

図23は、管理システム1による出力処理に係る動作シーケンスの一例を示す。

【0191】

以下に説明する動作シーケンスは、管理システム1が有する各サーバ又は装置の記憶装置に予め記憶されているプログラムに基づいて、主に各サーバ又は装置の制御装置により、各サーバ又は装置の各要素と協働して実行される。この動作シーケンスは、定期的に行われる。

【0192】

まず、端末装置10は、利用者が特許文献を検索するための検索画面を表示するための検索画面表示データの取得を要求するための検索画面表示データ要求信号をUIサーバ12に送信する(ステップS301)。

【0193】

UIサーバ12は、端末装置10から検索画面表示データ要求信号を受信した場合、検索画面表示データを端末装置10に送信する(ステップS302)。検索画面表示データは、HTML又はJava(登録商標)script等の公知のプログラミング言語によって生成される。

【0194】

端末装置10は、UIサーバ12から検索画面表示データを受信した場合、検索画面表示データに従って、検索画面を表示する(ステップS303)。端末装置10は、ウェブブラウザ等に検索画面を表示する。

【0195】

図24は、検索画面2400の一例を示す模式図である。

【0196】

図24に示す検索画面2400は、対象特許庁が日本国特許庁である場合の検索画面の一例である。図24に示すように、検索画面2400には、検索指定領域2410及び検索結果表示領域2430が含まれる。なお、UIサーバ12から受信する表示データには、検索結果表示領域2430を表示するためのデータが含まれておらず、ステップS303で表示される検索画面には、検索結果表示領域2430は表示されない。検索指定領域2410には、発行国指定ボックス2411、種別選択ボックス2412、審査対象指定ボックス2413、テーマ指定ボタン2414、公知日指定ボックス2415、検索式指定ボックス2416、画像指定ボックス2417、概念検索選択ボタン2418、機械翻訳文選択ボタン2419、推定分類選択ボタン2420及び検索ボタン2421等が含まれる。

【0197】

発行国指定ボックス2411は、検索対象の特許文献を発行した国を指定するためのボックスである。利用者は、検索対象の特許文献を発行した国として外国を指定する場合、さらにその国を指定することができる。種別選択ボックス2412は、検索対象の特許文献が全文献であるか公開された文献のみであるかを選択するためのボックスである。審査対象指定ボックス2413は、審査対象の特許文献を指定するためのボックスである。テーマ指定ボタン2414は、テーマコードを指定するためのボタンである。テーマ指定ボタン2414が押下されると、テーマコードを入力するためのボックスが表示される。公知日指定ボックス2415は、検索対象の特許文献の公知日を指定するためのボックスである。検索式指定ボックス2416は、検索式(キーワード、FI及び/又はFターム)を入力するためのボックスである。

【0198】

画像指定ボックス2417は、画像を入力するためのボックスである。なお、画像指定ボックス2417には、審査対象指定ボックス2413で指定された審査対象の特許文献に含まれる図面内の画像が選択可能に表示されてもよい。画像指定ボックス2417において画像が指定された場合、指定された画像と類似する画像を含む図面が記載された特許文献が検索される。概念検索選択ボタン2418は、概念検索を実行するか否かを選択するためのボタンである。機械翻訳文選択ボタン2419は、AIサーバ400により生成

10

20

30

40

50

された翻訳文を検索対象に含ませるか否かを選択するためのボタンである。推定分類選択ボタン2420は、AIサーバ400により生成された分類を検索対象に含ませるか否かを選択するためのボタンである。検索ボタン2421は、上記した各ボタン及びボックスで指定又は選択した条件で検索を実行させるためのボタンである。

【0199】

利用者により検索ボタン2421が押下された場合、端末装置10は、検索の実行を要求するための検索要求信号をゲートウェイサーバ11に送信する(ステップS304)。検索要求信号には、検索画面2400で指定された各条件(検索クエリ)、即ち利用者によって指定された、検索のための指定データが含まれる。

【0200】

ゲートウェイサーバ11は、端末装置10から検索要求信号を受信した場合、検索の実行を指示するための検索指示信号を複数の検索サーバ500に送信する(ステップS305)。ゲートウェイサーバ11は、検索要求信号に含まれる検索クエリを検索サーバ500が処理可能なフォーマット(SQLフォーマット等)に変換し、変換した検索クエリを含む検索指示信号を各検索サーバ500に送信する。なお、ゲートウェイサーバ11は、各端末装置10から受信した検索要求信号に含まれる検索クエリを受付キューに登録し、検索サーバ500の処理状況を踏まえて検索指示信号を送信する。これにより、ゲートウェイサーバ11は、ロードバランサとして機能し、検索サーバ500の負荷を平準化することができる。

【0201】

各検索サーバ500の検索部522は、第5通信装置501を介してゲートウェイサーバ11から検索指示信号を受信した場合、検索指示信号に含まれる検索クエリに従って、検索処理を実行する(ステップS306)。検索処理において、検索部522は、検索用データベース600に記憶された第3統合情報から、検索指示信号に含まれる検索クエリ(条件)を満たす特許文献、即ち利用者によって指定された指定データに対応する複数の特許文献を並べて表示するための第1表示データを生成する。検索処理の詳細については後述する。

【0202】

次に、検索部522は、検索処理において生成した第1表示データを、第5通信装置501を介してゲートウェイサーバ11に送信する(ステップS307)。

【0203】

ゲートウェイサーバ11は、各検索サーバ500から第1表示データを受信した場合、各検索サーバ500から受信した第1表示データを統合して、端末装置10に送信する(ステップS308)。

【0204】

端末装置10は、ゲートウェイサーバ11から統合された第1表示データを受信した場合、受信した第1表示データを記憶するとともに、第1表示データに従って、複数の特許文献を並べて表示する(ステップS309)。

【0205】

端末装置10は、ゲートウェイサーバ11から統合された検索結果を受信した場合、受信した第1表示データを記憶するとともに、第1表示データに従って、複数の特許文献を並べて表示する(ステップS309)。

【0206】

図24に示すように、ゲートウェイサーバ11から第1表示データを受信した場合、検索画面2400には、検索結果表示領域2430がさらに表示される。検索結果表示領域2430には、検索結果に示される特許文献毎に、各特許文献の文献番号2431、テーマコード2432、FI2433、公知日2434、発明の名称2435、チェックボックス2436等が含まれる。また、検索結果表示領域2430には、スクロールバー2437及び更新ボタン2438が含まれる。

【0207】

10

20

30

40

50

検索結果表示領域 2430 において、各特許文献は、検索処理において決定された順番に従って表示される。チェックボックス 2436 は、利用者が注目する特許文献を指定するためのボタンである。スクロールバー 2437 は、検索結果表示領域 2430 に一度に表示できない数の特許文献が含まれる場合、表示されていない特許文献を表示可能に、検索結果表示領域 2430 をスクロールさせるためのバーである。更新ボタン 2438 は、各特許文献を、チェックボックス 2436 で指定された特許文献との類似度に基づいて並べ替えて表示させるためのボタンである。

**【0208】**

利用者により更新ボタン 2438 が押下された場合、端末装置 10 は、特許文献の並べ替えを要求するための更新要求信号をゲートウェイサーバ 11 に送信する（ステップ S310）。更新要求信号には、チェックボックス 2436 で指定された特許文献、即ち端末装置 10 において利用者によって指定された特許文献を示す情報が含まれる。なお、更新要求信号には、検索結果表示領域 2430 において所定時間以上連続して表示された特許文献を示す情報が含まれてもよい。

**【0209】**

ゲートウェイサーバ 11 は、端末装置 10 から更新要求信号を受信した場合、特許文献の並べ替えを指示するための更新指示信号を複数の検索サーバ 500 に送信する（ステップ S305）。ゲートウェイサーバ 11 は、更新要求信号に含まれる特許文献を示す情報を検索サーバ 500 が処理可能なフォーマット（SQL フォーマット等）に変換し、変換した検索クエリを含む更新指示信号を各検索サーバ 500 に送信する。

**【0210】**

各検索サーバ 500 の検索部 522 は、第 5 通信装置 501 を介してゲートウェイサーバ 11 から更新指示信号を受信した場合、更新指示信号に含まれる特許文献を示す情報に従って、更新処理を実行する（ステップ S312）。更新処理において、検索部 522 は、第 1 表示データにより表示される特許文献を、更新指示信号において指定された特許文献との類似度に基づいて並べ替えた第 2 表示データを生成する。更新処理の詳細については後述する。

**【0211】**

次に、検索部 522 は、検索処理において生成した第 2 表示データを、第 5 通信装置 501 を介してゲートウェイサーバ 11 に送信する（ステップ S313）。

**【0212】**

ゲートウェイサーバ 11 は、各検索サーバ 500 から第 2 表示データを受信した場合、各検索サーバ 500 から受信した第 2 表示データを統合して、端末装置 10 に送信する（ステップ S314）。なお、更新処理は、一つの検索サーバ 500 により実行されてもよい。その場合、ゲートウェイサーバ 11 は、一つの検索サーバ 500 から第 2 表示データを端末装置 10 に送信する。

**【0213】**

端末装置 10 は、ゲートウェイサーバ 11 から第 2 表示データを受信した場合、受信した第 2 表示データを記憶するとともに、第 2 表示データに従って、第 1 表示データにより表示される特許文献を並べ替えて表示する（ステップ S315）。以上により、出力処理に係る動作シーケンスは終了する。

**【0214】**

なお、検索サーバ 500 は、受信した検索指示信号に含まれる検索クエリに示される審査対象の特許文献、テーマコード（技術分野の大分類）及び検索式に含まれる F I、F ターム（技術分野の小分類）等を A I サーバ 400 にフィードバックさせてもよい。その場合、検索サーバ 500 は、その審査対象の特許文献と、技術分野の大分類と、技術分野の小分類とを A I サーバ 400 に送信する。A I サーバ 400 は、受信した特許文献と、技術分野の大分類と、技術分野の小分類とを用いて、技術分野を特定するための学習モデルを更新する。これにより、管理システム 1 は、学習モデルにより特定される技術分野の精度を継続的に向上させることが可能となる。

10

20

30

40

50

## 【0215】

図25は、検索サーバ500における検索処理の動作の例を示すフローチャートである。

## 【0216】

図25に示す検索処理は、図23に示す出力処理のステップS306で実行される。

## 【0217】

まず、検索部522は、受信した検索指示信号に含まれる検索クエリ(条件)において、概念検索を実行することが指定されているか否かを判定する(ステップS401)。

## 【0218】

概念検索を実行することが指定されていない場合、検索部522は、検索用データベース600に記憶された第3統合情報を参照して、検索指示信号に含まれる検索クエリ(条件)を満たす特許文献を抽出し(ステップS402)、処理をステップS408へ移行する。検索部522は、検索用データベース600に記憶された特許文献の中から、検索クエリで指定された発行国、公知日、テーマコード、FI、Fターム及び/又は各キーワードが、第3統合情報において対応する各データ項目と合致する特許文献を抽出する。検索用データベース600には、一つの特許文献に関する複数の情報が第3統合情報としてまとめて記憶される。検索サーバ500は、ユーザが複数の情報を指定して検索を行った場合でも、指定された複数の情報をまとめて検索できるため、検索処理の検索時間を短縮し、処理負荷を軽減させることができる。

## 【0219】

なお、検索クエリにおいて、AIサーバ400により生成された翻訳文又は分類を検索対象に含ませることが指定されている場合、検索部522は、AIサーバ400により生成された翻訳文又は分類を対象として、指定された検索式を満たす特許文献を抽出する。検索用データベース600には、対象特許庁における使用言語と異なる言語で記載された特許文献について、その使用言語で記載された翻訳文が記憶される。検索部522は、検索クエリに含まれるキーワードが、対象特許庁における使用言語で指定されている場合、その使用言語と異なる言語で記載された特許文献については、その使用言語による翻訳文に対して検索を行う。そのため、利用者は、各特許文献における言語の違いを意識することなく、様々な言語で記載された特許文献を一括して効率良く検索することができ、管理システム1は、利用者の利便性を向上させることができる。

## 【0220】

なお、検索部522は、検索クエリに含まれるキーワードが、対象特許庁における使用言語以外の言語で指定されている場合、その言語で記載された特許文献に対して検索を行ってもよい。これにより、検索部522は、より精度良く検索を実行することができる。

## 【0221】

一方、概念検索を実行することが指定されている場合、検索部522は、検索指示信号に含まれる検索クエリにおいて指定されたキーワード又は画像、即ち利用者によって指定された指定データについて複数の第1特徴ベクトルを生成する(ステップS403)。検索部522は、図21に示した検索情報生成処理のステップS204及びS205と同様にして、検索クエリにおいて指定されたキーワード又は画像の第1特徴ベクトルを生成する。

## 【0222】

次に、検索部522は、比較対象の第1特徴ベクトルを選択する(ステップS404)。検索部522は、ステップS404の処理が最初に実行される時は、複数の第1特徴ベクトルの内、次元数が最大である第1特徴ベクトルを比較対象の第1特徴ベクトルとして選択する。検索部522は、ステップS404の処理が二回目以降に実行される時は、前回選択された第1特徴ベクトルの次に次元数が大きい第1特徴ベクトルを比較対象の第1特徴ベクトルとして選択する。

## 【0223】

次に、検索部522は、検索用データベース600に記憶された第3統合情報を参照し

て、検索指示信号に含まれる検索クエリ（条件）を満たす特許文献を抽出する（ステップ S 4 0 5）。検索部 5 2 2 は、検索用データベース 6 0 0 に記憶された特許文献の中から、検索クエリで指定された発行国、公知日、テーマコード、各キーワード、F I、F ターム及び/又は比較対象の第 1 特徴ベクトルが、第 3 統合情報において対応する各データ項目と合致する特許文献を抽出する。

#### 【 0 2 2 4 】

一般に、情報処理装置は、二つの特徴ベクトルが対応するか否かを判定する場合、二つの特徴ベクトルの類似度（例えばコサイン類似度等）を算出し、類似度が閾値以上であるか否かにより、二つの特徴ベクトルが対応するか否かを判定する。そのため、情報処理装置は、発行国、公知日、テーマコード、各キーワード、F I、F タームが合致する特許文献を抽出してから、抽出した特許文献毎に特徴ベクトルの類似度を算出し、類似度が閾値以上である特許文献を特定する必要がある。

#### 【 0 2 2 5 】

一方、第 1 特徴ベクトルは、L S H を用いて各特許文献の第 2 特徴ベクトルを変換したハッシュ値を要素とする特徴ベクトルであり、その次元数の等間隔な超平面集合で第 2 特徴ベクトルの特徴量空間を分割した各部分空間（バケツ）を表す。検索サーバ 5 0 0 は、二つの第 1 特徴ベクトルが対応するか否かを判定する場合、二つの第 1 特徴ベクトルが表す部分空間が一致するか否かを判定することにより、二つの第 1 特徴ベクトルが対応するか否かを高精度に判定することができる。即ち、検索サーバ 5 0 0 は、第 1 特徴ベクトルの各要素（ハッシュ値）が一致するか否かを判定するだけで、二つの第 1 特徴ベクトルが対応するか否かを高精度に判定することができる。そのため、検索サーバ 5 0 0 は、第 1 特徴ベクトル同士の比較を文字列同士の比較と同様に扱うことができ、発行国、公知日、テーマコード、各キーワード、F I、F タームの照合と、第 1 特徴ベクトルの照合とをまとめて実行することができる。したがって、検索サーバ 5 0 0 は、検索処理の処理負荷を軽減させることができ、多数の特許文献を高速に検索することができる。

#### 【 0 2 2 6 】

次に、検索部 5 2 2 は、抽出した特許文献毎に、評価値を算出する（ステップ S 4 0 6）。例えば、検索部 5 2 2 は、各特許文献の評価値の初期値を 0 とし、各特許文献を抽出するたびに、各特許文献を抽出した際の第 1 特徴ベクトルの次元数を評価値に加算する。次元数が大きいほど第 1 特徴ベクトルが表す情報量は大きく、抽出した特許文献が、検索クエリにおいて指定されたキーワード又は画像に対応している可能性が高い。そのため、検索部 5 2 2 は、次元数が大きいほど評価値を大きくすることにより、検索クエリにおいて指定されたキーワード又は画像に対応する特許文献を精度良く抽出することができる。

#### 【 0 2 2 7 】

次に、検索部 5 2 2 は、評価値が基準値以上である特許文献の数が所定数以上であるか否かと、比較対象の第 1 特徴ベクトルの次元数が最小の次元数であるか否かと、を判定する（ステップ S 4 0 7）。基準値及び所定数は、事前に設定される。評価値が基準値以上である特許文献の数が所定数未満であり且つ比較対象の第 1 特徴ベクトルの次元数が最小の次元数でない場合、検索部 5 2 2 は、処理をステップ S 4 0 4 へ戻し、ステップ S 4 0 4 ~ S 4 0 7 の処理を繰り返す。このように、検索部 5 2 2 は、第 1 特徴ベクトルの次元数の降順に、指定データについて生成された第 1 特徴ベクトルと、各特許文献について生成された第 1 特徴ベクトルとを比較することにより、指定データに対応する特許文献を抽出する。検索部 5 2 2 は、情報量が大きい順に第 1 特徴ベクトルを比較することにより、指定データに近似する特許文献を精度良く抽出することができる。また、検索部 5 2 2 は、十分な数の特許文献を抽出した時点で検索処理を終了することができ、検索処理の処理時間を短縮させることができる。また、検索部 5 2 2 は、最大でも、第 1 特徴ベクトルの最大次元数分だけステップ S 4 0 4 ~ S 4 0 7 の処理を繰り返すことにより、検索処理を完了させることができ、検索処理の処理時間が増大することを抑制できる。

#### 【 0 2 2 8 】

一方、評価値が基準値以上である特許文献の数が所定数以上である場合、又は、比較対

象の第1特徴ベクトルの次元数が最小の次元数である場合、検索部522は、抽出した各特許文献を並べて表示するための第1表示データを生成し(ステップS408)、一連のステップを終了する。これにより、検索部522は、利用者によって指定された指定データに対応する複数の特許文献を並べて表示するための第1表示データを生成する。検索部522は、抽出した各特許文献を、公知日の順に従って又はランダムに並べて表示するように第1表示データを生成する。なお、検索部522は、抽出した特許文献毎に、指定データとの類似度(各特許文献の特徴ベクトルと指定データの特徴ベクトルのコサイン類似度又はユークリッド距離等)を算出し、抽出した各特許文献を類似度が高い順に並べて表示してもよい。

#### 【0229】

このように、検索部522は、利用者からの要求に従って、第3統合情報を用いて複数の特許文献を検索する。なお、検索部522は、第1特徴ベクトルの次元数の昇順に、指定データについて生成された第1特徴ベクトルと、各特許文献について生成された第1特徴ベクトルとを比較することにより、指定データに対応する特許文献を抽出してもよい。

#### 【0230】

その場合、検索部522は、ステップS404の処理が最初に実行される時は、複数の第1特徴ベクトルの内、次元数が最小である第1特徴ベクトルを比較対象の第1特徴ベクトルとして選択する。検索部522は、ステップS404の処理が二回目以降に実行される時は前回選択された第1特徴ベクトルの次に次元数が小さい第1特徴ベクトルを比較対象の第1特徴ベクトルとして選択する。また、ステップS406の処理は省略され、ステップS407において、検索部522は、抽出した特許文献の数が所定範囲内であるか否かと、比較対象の第1特徴ベクトルの次元数が最大の次元数であるか否かと、を判定する。抽出した特許文献の数が所定範囲内でない且つ比較対象の第1特徴ベクトルの次元数が最大の次元数でない場合、検索部522は、処理をステップS404へ戻し、ステップS404～S407の処理を繰り返す。一方、抽出した特許文献の数が所定範囲内である場合、又は、比較対象の第1特徴ベクトルの次元数が最大の次元数である場合、ステップS408において、検索部522は、抽出した各特許文献を並べて表示するための第1表示データを生成し、一連のステップを終了する。

#### 【0231】

利用者によって指定された指定データに対応する特許文献の数が少ない場合、検索部522は、情報量が小さい順に第1特徴ベクトルを比較することにより、より早期に適切な数の特許文献を抽出することができ、検索処理の処理時間を短縮させることができる。

#### 【0232】

また、検索部522が生成する第1特徴ベクトルは、LSHを用いて各特許文献の第2特徴ベクトルを変換したハッシュ値を要素とする特徴ベクトルに限定されない。複数の第1特徴ベクトルは、それぞれ次元数が異なる特徴ベクトルであればよく、TF-IDF、Bag of Words又はBM25等を各要素とする特徴ベクトルでもよい。

#### 【0233】

図26は、検索サーバ500における更新処理の動作の例を示すフローチャートである。

#### 【0234】

図26に示す更新処理は、図23に示す出力処理のステップS312で実行される。

#### 【0235】

まず、検索部522は、受信した更新指示信号に含まれる特許文献を示す情報から、端末装置10において利用者によって指定された特許文献又は所定時間以上連続して表示された特許文献を特定する(ステップS501)。

#### 【0236】

次に、検索部522は、図25に示した検索処理のステップS408で生成した第1表示データにより表示される各特許文献と、ステップS501で特定した特許文献との類似度を算出する(ステップS502)。検索部522は、類似度として、各特許文献の第1

10

20

30

40

50

特徴ベクトルのコサイン類似度又はユークリッド距離等を算出する。

【0237】

次に、検索部522は、算出した類似度に基づいて、第1表示データにより表示される特許文献を並べ替えた第2表示データを生成し(ステップS503)、一連のステップを終了する。検索部522は、第1表示データにより表示される特許文献を、類似度が高い順に並べて表示するように第2表示データを生成する。これにより、利用者は、注目する特許文献に類似する特許文献から順番に確認することが可能となり、管理システム1は、利用者の利便性を向上させることができる。なお、検索部522は、第1表示データにより表示される特許文献の内、先頭の特許文献から、更新指示信号で指定された特許文献までは元の順番で並べつつ、更新指示信号で指定された特許文献から、最後の特許文献までのみを類似度に基づいて並べ替えてもよい。これにより、利用者が既に確認した特許文献については現状の順番のままとなり、利用者は、既に確認した特許文献を重複して確認する必要がなくなり、管理システム1は、利用者の利便性を向上させることができる。

10

【0238】

以上説明してきたように、管理サーバ300は、照会サーバ100から取得した形式情報と、書誌サーバ200から取得した書誌情報とを統合した第1統合情報を生成してから、AIサーバ400から受信した検索情報をさらに統合して第1統合情報を生成する。これにより、管理サーバ300は、第2統合情報を短時間に生成することが可能となり、特許文献に関する情報を効率良く管理することが可能となった。

20

【0239】

また、管理システム1は、各サーバにおける処理負荷及びネットワーク負荷を抑制することが可能となり、小規模なサーバ群によってビッグデータを検索サーバに蓄積することが可能となり、システム構築費用を低減させることが可能となった。

【0240】

また、管理システム1では、短期間で開発を見直す手法を取り且つ開発内容に柔軟性を持たせることが可能なアジャイル開発手法により、利用者からのフィードバックを踏まえて短期間でPDCA(Plan Do Check Act)サイクルを循環させることが可能となった。

【0241】

また、各特許文献には、テキスト、技術分野の分類及び画像等の様々な情報が含まれる。仮に、テキスト、技術分野の分類及び画像がそれぞれ異なる検索サーバにまたがって検索される場合、各検索サーバに大きな負荷がかかる。例えば5000万件の特許文献を対象として検索が行われ、ある検索キーによって1つの検索サーバで3000万件がヒットし、他の検索サーバで2000万件がヒットし、さらに他の検索サーバで1000万件がヒットした場合を想定する。その場合、トータルとして(3000万×2000万×1000万)回だけ、各サーバでヒットした特許文献を突合せさせる必要があり、検索サーバの負荷は膨大となる。また、各サーバでヒットした特許文献を突合せさせる処理の負荷量は、検索サーバの数に応じて指数関数的に増大していく。

30

【0242】

管理システム1では、一つの特許文献に関する情報が複数の検索サーバに分割されて記憶されるのではなく、一つの特許文献に関する全ての情報が一つの検索サーバ500に記憶される。そのため、ユーザが複数の情報を指定して検索を行った場合でも、各検索サーバ500が、指定された複数の情報をまとめて検索するため、効率良く検索することができ、各検索サーバ500におけるトータルの処理負荷を軽減させることができる。これにより、管理システム1は、大規模なサーバ群を構築することなく円滑に検索を行うことができ、サーバの導入にかかる各種設定、サーバの運用及び保守等にかかる人件費と、設置スペースに係る費用とを抑制することができる。また、管理システム1は、検索条件に上限等を設定することなく円滑に検索することができ、検索漏れの発生等を抑制することができる。

40

【0243】

以上、好適な実施形態について説明してきたが、実施形態はこれらに限定されない。例

50



えば、管理システムが有する各サーバの機能分担は、図1に示す管理システム1の例に限られず、各サーバの各部を何れのサーバに配置するかは適宜変更可能である。または、管理システム1が有する全てのサーバを一つのサーバで構成してもよい。

【0244】

また、管理システム1が管理する文献は、特許文献に限定されず、分類に紐づく文献であれば、どのような文献でもよく、例えば論文、ニュース記事等の非特許文献でもよい。

【0245】

また、AIサーバ400が使用する各学習モデルは、AIサーバ400により生成されて第4記憶装置410に記憶されるのではなく、外部のサーバに生成されて外部のサーバに記憶されていてもよい。その場合、検索情報生成部421は、各学習モデルに入力する情報を外部のサーバに送信し、各学習モデルからの出力値を外部のサーバから受信してもよい。

【0246】

以下、図25のステップS403～S407で説明した概念検索の他の実施形態について説明する。

【0247】

検索サーバ500は、第5記憶装置510に記憶された特許文献から、直近の所定期間（例えば数年間）に登録された特許文献を抽出し、各特許文献の全文において出現頻度が高い順に上位所定数の用語を特徴語として抽出する。そして、検索サーバ500は、第5記憶装置510に記憶された特許文献毎に、全文内の各特徴語の出現回数を要素とする第3特徴ベクトルと、特許請求の範囲内の各特徴語の出現回数を要素とする第4特徴ベクトルとを生成しておく。

【0248】

検索部522は、概念検索を実行する場合、審査対象の特許文献の全文内の各特徴語の出現回数を要素とする第3特徴ベクトルと、審査対象の特許文献の特許請求の範囲内の各特徴語の出現回数を要素とする第4特徴ベクトルとを生成する。次に、検索部522は、審査対象の特許文献の第3特徴ベクトルと、第5記憶装置510に記憶された各特許文献の第3特徴ベクトルとの第1類似度を算出する。第1類似度は、例えば正規化相互相関値である。次に、検索部522は、第5記憶装置510に記憶された特許文献の内、第1類似度が高い順に上位第1所定数の特許文献を、審査対象の特許文献に類似する特許文献として抽出する。次に、検索部522は、審査対象の特許文献の第4特徴ベクトルと、抽出した各特許文献の第4特徴ベクトルとの第2類似度を算出し、抽出した各特許文献を示す情報を第2類似度が高い順に並べて、検索結果として送信する。第2類似度は、例えば正規化相互相関値である。なお、検索部522は、抽出した各特許文献の内、第2類似度が高い順に上位第2所定数の特許文献を示す情報のみを検索結果として送信してもよい。一方、端末装置10は、検索結果に示される各特許文献を示す情報を、検索結果に示される順に並べて表示する。これにより、利用者は、各特許文献を類似度が高い順に効率良く参照することができる。

【0249】

なお、検索部522は、審査対象の特許文献の特許請求の範囲に記載された各文を読点等で区切って複数の構成要素に分割し、利用者に各構成要素を提示し、提示した構成要素の中から、絞り込み検索に用いられる一又は複数の構成要素を選択させてもよい。構成要素が所定文字数未満である場合、検索部522は、その構成要素をその構成要素に続く構成要素と連結させてもよい。また、検索部522は、利用者に、選択された構成要素と、その構成要素に含まれる各単語とを提示し、提示した単語の中から、絞り込み検索に用いられる一又は複数の単語をさらに選択させてもよい。

【0250】

その場合、検索部522は、選択された各構成要素について、構成要素に含まれる単語及びそのシソーラスを含むことを条件とする検索式を作成し、審査対象の特許文献に類似する特許文献として抽出した各特許文献が、作成した検索式を満たす度合いを算出する。

検索部 5 2 2 は、抽出した各特許文献を示す情報を、算出した度合いが高い順に並べて、検索結果として送信する。なお、検索部 5 2 2 は、抽出した各特許文献の内、算出した度合いが高い順に上位第 3 所定数の特許文献を示す情報のみを検索結果として送信してもよい。また、検索部 5 2 2 は、検索結果において、各特許文献を示す情報に、各構成要素に対応する検索式を満たすか否かを示す情報を関連付けてもよい。端末装置 1 0 は、検索結果に示される各特許文献を示す情報を、各構成要素に対応する検索式を満たすか否かを示す情報とともに表示する。これにより、利用者は、所望の条件を満たす特許文献を効率良く参照することができる。

【 0 2 5 1 】

また、検索部 5 2 2 は、概念検索で抽出された各特許文献に含まれる各図面を、図 2 4 の検索画面 2 4 0 0 の画像指定ボックス 2 4 1 7 で指定された画像との一致度合いに基づいて、順位付けしてもよい。

【 0 2 5 2 】

その場合、検索部 5 2 2 は、二つの画像が入力された場合に、その二つの画像の一致度合いを出力するように事前学習された学習モデルを用いて、概念検索で抽出された特許文献に含まれる図面と、指定された画像との一致度合いを取得する。この学習モデルは、複数の学習用画像を用いて、例えばディープラーニングを用いて学習される。ニューラルネットワークである学習モデルは、入力層、中間層及び出力層から構成される多層構造を有する。入力層の各ノードには、特徴量として、二つの画像からそれぞれ抽出された対象物の角又は交点等の特徴点の数又は分布等に関する各情報が入力される。中間層の各ノードには、入力層の各ノードから出力された値の、所定の重みに基づく重み付け和が入力される。出力層の各ノードには、中間層の各ノードから出力された値の、所定の重みに基づく重み付け和が入力される。出力層は、例えば、入力された値をそのまま出力する。各重みは、二つの画像が類似している程、出力層によって出力される値が大きくなるように設定される。重みの設定には、例えば、誤差逆伝搬法などの公知の手法が使用される。検索部 5 2 2 は、概念検索で抽出された各特許文献に含まれる各図面に含まれる画像と、指定された画像とから特徴量を算出し、算出した特徴量を学習モデルに入力し、その一致度合いを取得する。

【 0 2 5 3 】

検索部 5 2 2 は、検索結果において、各特許文献に含まれる各図面を示す情報に、各図面について算出した一致度合いを関連付ける。端末装置 1 0 は、検索結果に示される各特許文献に含まれる各図面を、その一致度合いが高い順に表示する。これにより、利用者は、所望の画像を含む特許文献の図面を効率良く参照することができる。また、この場合、端末装置 1 0 は、検索結果として表示された図面の内の何れかの図面が利用者により指定された場合、指定された図面に対応する記載又は図面の説明をテキスト表示してもよい。また、端末装置 1 0 は、利用者により指定された図面に類似する順に、各図面をソートしてもよい。

【 0 2 5 4 】

図 2 7 は、他の実施形態に係る管理システム 2 による処理の一例を説明するための模式図である。

【 0 2 5 5 】

管理システム 2 では、情報処理装置 1 5 は、所定期間毎に、所定の特許庁データベースから、新たに出願又は登録された特許文献を収集し、照会サーバ 1 0 0、書誌サーバ 2 0 0 及び A I サーバ 4 0 0 に配信する。照会サーバ 1 0 0 は、所定期間に収集した新規の各特許文献から形式情報を取得する。書誌サーバ 2 0 0 は、各特許文献から書誌情報を抽出する。管理サーバ 3 0 0 は、各特許文献について、形式情報と書誌情報を統合した第 1 統合情報を生成する。一方、A I サーバ 4 0 0 は、管理サーバ 3 0 0 が、第 1 統合情報を生成している間に、各特許文献の書誌情報又は内容情報に基づいて、各特許文献には記載されず且つ検索に利用される検索情報を生成する。そして、管理サーバ 3 0 0 は、第 1 統合情報が完成した後に、各特許文献について、第 1 統合情報と検索情報を統合した第 2 統合

情報を生成する。この第2統合情報は、検索用データベースに登録を行うために使用される。その他の点については、管理システム2は、管理システム1と同様に動作する。管理システム2も、特許文献に関する情報を効率良く管理することが可能となる。

【符号の説明】

【0256】

1 管理システム

- 100 照会サーバ
- 200 書誌サーバ
- 300 管理サーバ
- 400 AIサーバ
- 500 検索サーバ
- 600 検索用データベース

【要約】 (修正有)

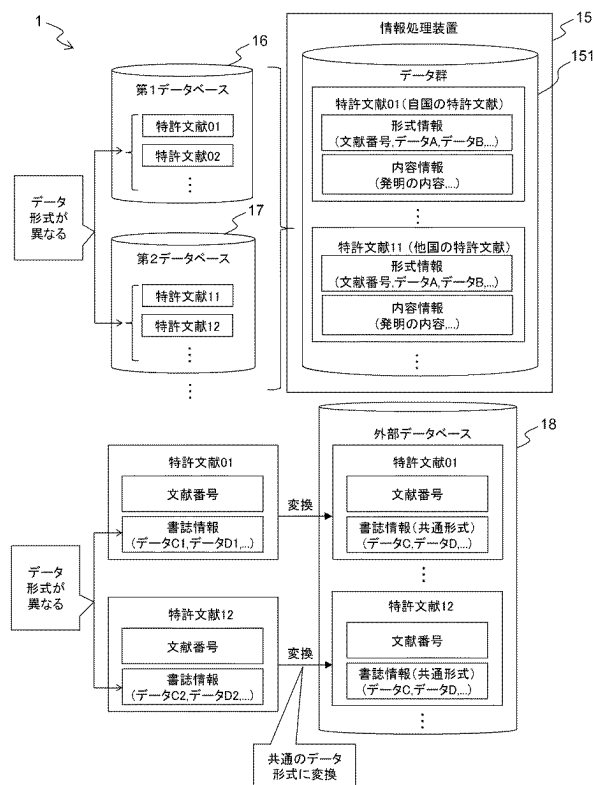
【課題】特許文献に関する情報を効率良く管理することができる管理システム及び管理方法を提供する。

【解決手段】管理システム1は、照会サーバ100と、書誌サーバ200と、管理サーバ300と、AIサーバ400と、を有する。照会サーバ100は、複数の特許文献について形式情報と内容情報とが含まれるデータ群151から各特許文献について形式情報を取得する。書誌サーバ200は、外部データベース18から、各特許文献について、文献番号と共通のデータ形式に変換された書誌情報とを取得する。管理サーバ300は、照会サーバ100が形式情報を取得した各特許文献について、形式情報及び書誌情報を統合した第1統合情報を生成する。AIサーバ400は、各特許文献の分類情報を生成する。管理サーバ300は、第1統合情報が完成した後に、各特許文献について第1統合情報及び分類情報を統合した第2統合情報を生成する。

【選択図】図4

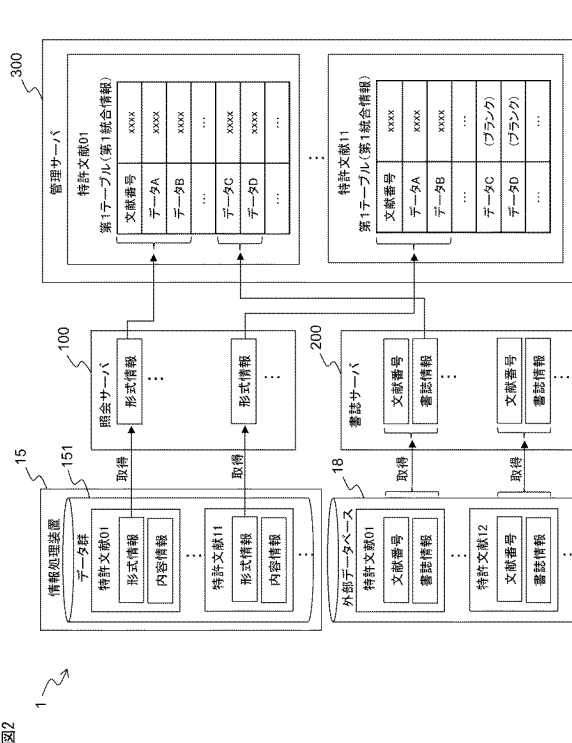
【図1】

図1



【図2】

図2



10

20

【図3】

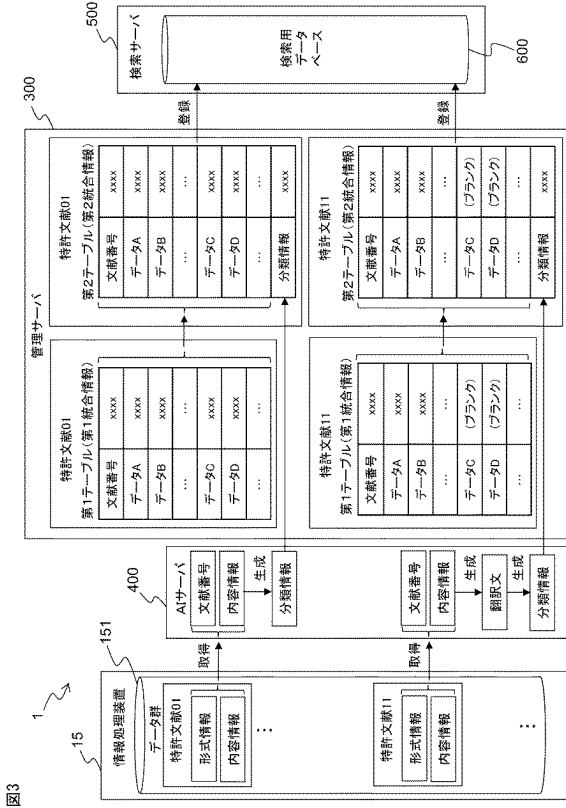


図3

【図4】

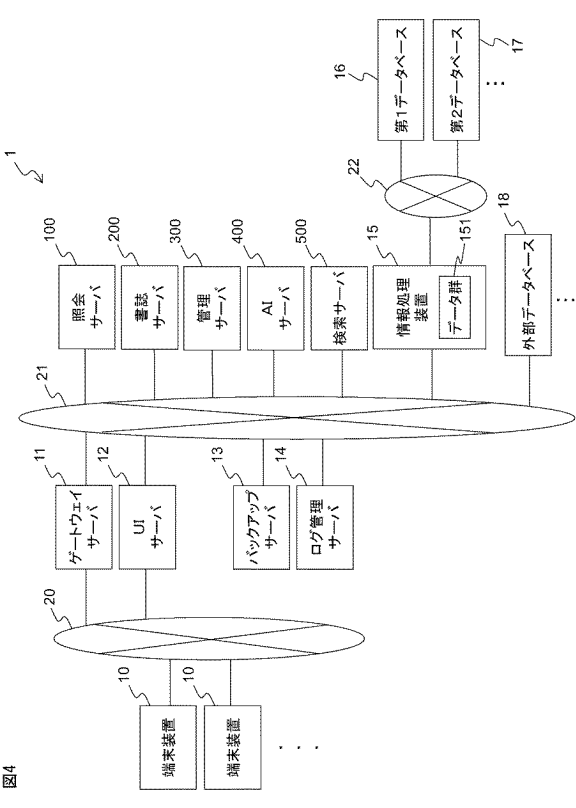
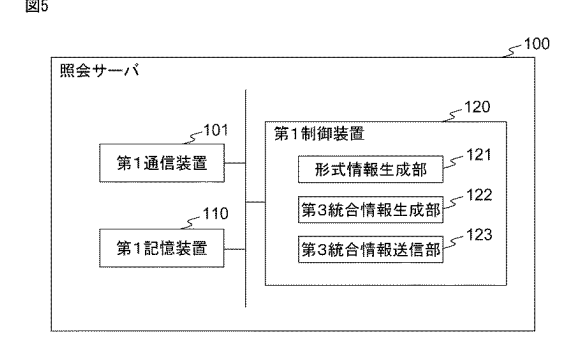


図4

【図5】



【図6】

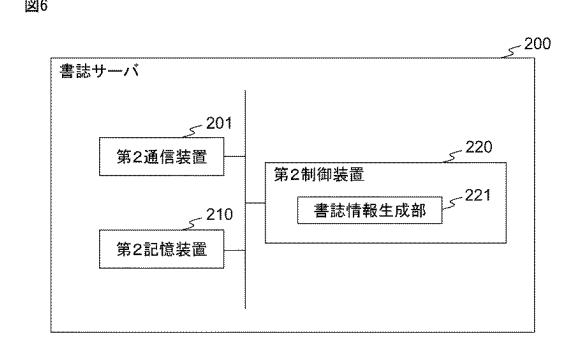


図6

【図7】

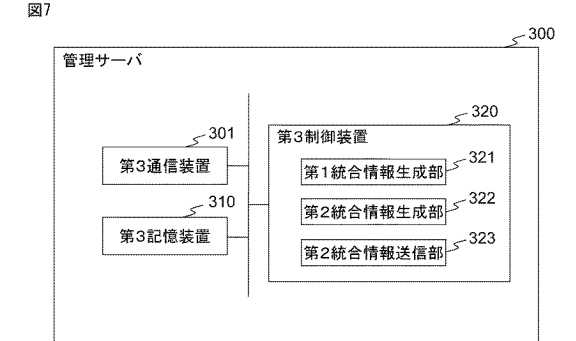
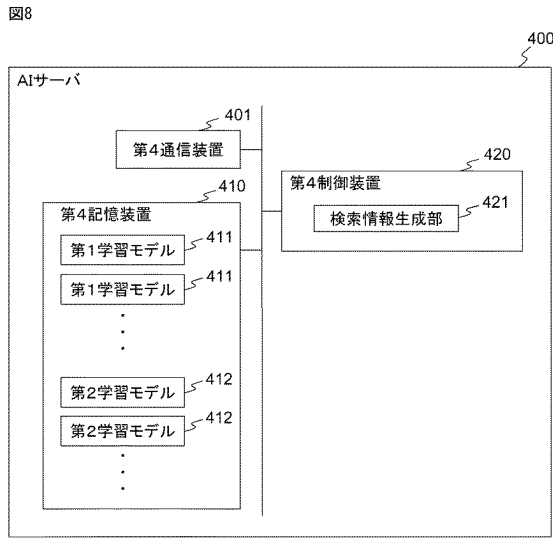


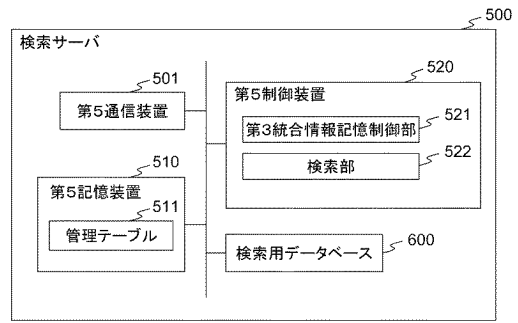
図7

【 図 8 】



【 図 9 】

図9



【 図 1 0 】

図10

フィールド名	データ型	インデックス	レコード1	レコード2
文献番号	String	-	JP2020000000A	WO2020000000A
発行機関	String	-	JPO	WIPO
言語	String	-	JA	EN
文献種別	String	-	A	A
出願番号	String	-	JP2019000000	PCT/WO2019/000000
発行日	Integer	-	20200630	20200630
出願日	Integer	-	20190101	20190101
テーマコード	String	"5A000"/"5A001"	"5A000"/"5A001"	-
Fターム	String	"AA01"/"AA02", "AA03"	"AA01"/"AA02", "AA03"	-
FI	String	"A01A1/00,100", "A01A1/00,110", "A01A1/00,110"	"A01A1/00,100", "A01A1/00,110"	-
IPC	String	"A01A1/00", "A01A1/01"	"A01A1/00", "A01A1/01"	"A01A1/02", "A01A1/03"
出願人名	String	特許太郎株式会社 /Patent Inc.	特許太郎株式会社	Patent Inc.
発明者名	String	特許太郎 /Patent Taro	特許太郎	Patent Taro
...	...	...	...	...

【 図 1 1 】

図11

フィールド名	データ型	インデックス	レコード1	レコード2
テーマコード_機械	String	"5A002"/"5A003"	-	"5A002"/"5A003"
Fターム_機械	String	"AA04"/"AA05", "AA06"	-	"AA04"/"AA05", "AA06"
FI_機械	String	"A01A1/00,120", "A01A1/00,130"	-	"A01A1/00,120", "A01A1/00,130"
機械翻訳文_1	String	発明,...	-	この発明は...
翻訳方式_1	String	-	-	SMT_model_1
機械翻訳文_2	String	発明,...	-	この発明は...
翻訳方式_2	String	-	-	NMT_model_2
...	...	...	...	...
機械翻訳文_n	String	発明,...	-	この発明は...
翻訳方式_n	String	-	-	NMT_model_n
図面のメタデータ	Binary	-	00010001, 01110001,...	10011001, 11110101,...
第1特徴ベクトル	Binary	(1),(1,0),...	(1),(1,0),...	(1),(1,0),...
...	...	...	...	...

【図 1 2】

ファイル名	データ型	インデックス	レコード1	レコード2
発明の名称	String	発明 A/ Invention A	発明 A	Invention A
抄録	String	発明, .../ invention, ...	この発明は...	This invention is...
請求の範囲	String	請求の範囲, .../ Claim, ...	請求の範囲 1 は...	Claim 1 is...
詳細な説明	String	実施の形態, .../ embodiment, ...	実施の形態は...	The embodiment of...
全テキスト	String	発明, .../ invention, ...	この発明は...	This invention is...
...	...	...	...	...

図12

【図 1 3】

ファイル名	データ型	インデックス	レコード1	レコード2
フォーマットID	String	-	123456	123457
代表文献フラグ	Integer	-	1	1
...	...	...	...	...
更新日	Integer	-	20210101	20210105
格納アドレス	String	-	/home/doc/jp/...	/home/doc/wo/...
ファイル名	String	-	JP2020000000A.x nml	WO2020000000A. xml
検索サーバ名	String	-	kensaku1.jpo.go.jp	kensaku2.jpo.go.jp
照会サーバ名	String	-	shokai1.jpo.go.jp	shokai2.jpo.go.jp
...	...	...	...	...

図13

【図 1 4】

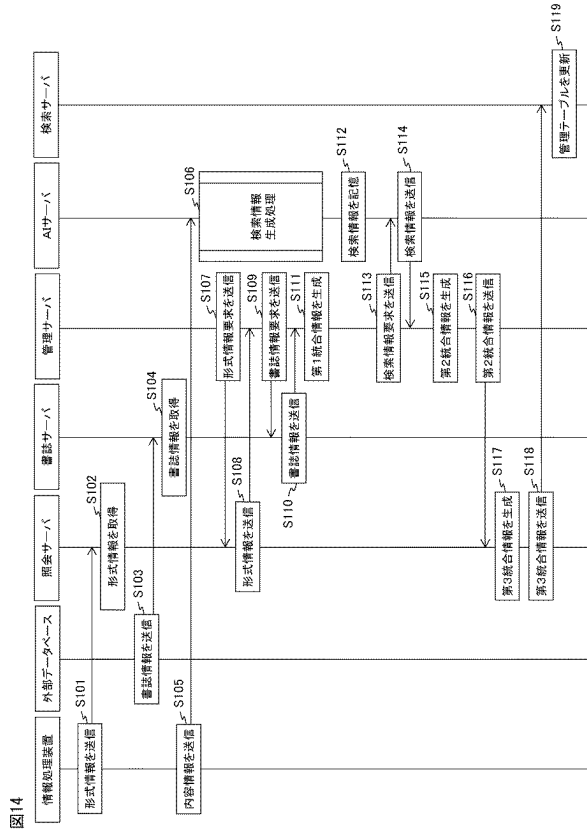


図14

【図 1 5】

ファイル名	データ型	レコード1	レコード2
発行機関	String	JPO	WIPO
文献番号	String	JP2020000000A	WO2020000000A
文献種別	String	A	A
格納アドレス	String	/var/www/bunken/data/ text/JP/2020/000/000/ JP2020000000A.xml	/var/www/bunken/data/text/wo/ 2020/000/000/ WO2020000000A.xml
ファイル名	String	ja	en
言語	String	ja	en
更新日	Integer	20210101	20210101

図15

【 図 1 6 】

フィールド名	データ型	レコード1	レコード2
発行機関	String	JPO	WIPO
文献番号	String	JP20200000000A	WO20200000000A
文献種別	String	A	A
出願番号	String	JP2019000000	PCT/WO2019/000000
発行日	Integer	20200630	20200630
出願日	Integer	20190101	20190101
FI	String	"A01A1/00,100", "A01A1/00,110"	-
テーマコード	String	"5A000","5A001"	-
Fターム	String	"AA01","AA02","AA03"	-
IPC	String	"A01A1/00","A01A1/01"	"A01A1/02","A01A1/03"
出願人名	String	特許太郎株式会社	Patent Inc.
発明者名	String	特許太郎	Patent Taro
更新日	Integer	20210101	20210101

図16

【 図 1 7 】

フィールド名	データ型	レコード1	レコード2
発行機関	String	JPO	WIPO
文献番号	String	JP20200000000A	WO20200000000A
文献種別	String	A	A
格納アドレス	String	/var/www/bunken/data/ text/JP/2020/000/000/ JP20200000000A.xml	/var/www/bunken/data/text/WO/ 2020/000/000/ WO20200000000A.xml
ファイル名	String	JP20200000000A.xml	WO20200000000A.xml
言語	String	ja	en
更新日	Integer	20210101	20210101
出願番号	String	JP2019000000	PCT/WO2019/000000
発行日	Integer	20200630	20200630
出願日	Integer	20190101	20190101
FI	String	"A01A1/00,100", "A01A1/00,110"	-
テーマコード	String	"5A000","5A001"	-
Fターム	String	"AA01","AA02","AA03"	-
IPC	String	"A01A1/00","A01A1/01"	"A01A1/02","A01A1/03"
出願人名	String	特許太郎株式会社	Patent Inc.
発明者名	String	特許太郎	Patent Taro
更新日	Integer	20210101	20210101

図17

【 図 1 8 】

フィールド名	データ型	レコード1	レコード2
文献番号	String	JP20200000000A	WO20200000000A
テーマコード_機械	String	-	"5A002","5A003"
Fターム_機械	String	-	"AA04","AA05","AA06"
FI_機械	String	-	"A01A1/00,120", "A01A1/00,130"
機械翻訳文_1	String	-	この発明は...
翻訳方式_1	String	-	SMT_model_1
機械翻訳文_2	String	-	この発明は...
翻訳方式_2	String	-	NMT_model_2
...	...	...	...
機械翻訳文_n	String	-	この発明は...
翻訳方式_n	String	-	NMT_model_n
図面のメタデータ	Binary	00010001,01110001..... 11110101.....	10011001,11110101..... (1),(1,0),(1,0,1,0),...
第1特徴ベクトル	Binary	(1),(1,0),(1,0,1,0),...	(1),(1,0),(1,0,1,0),...

図18

【 図 1 9 】

図19

フィールド名	データ型	レコード例 1	レコード例 2
発行機関	String	JPO	WIPO
文献番号	String	JP20200000000A	WO20200000000A
文献種別	String	A	A
格納場所	String	/var/www/bunken/data/ text/JP/2020/000/000/ JP20200000000A.xml	/var/www/bunken/data/ text/WO/2020/000/000/ WO20200000000A.xml
ファイル名	String	JP20200000000A.xml	WO20200000000A.xml
言語	String	ja	en
更新日	Integer	20210101	20210101
出願番号	String	JP2019000000	PCT/WO2019/000000
発行日	Integer	20200630	20200630
出願日	Integer	20190101	20190101
FI	String	"A01A1/00,100","A01A1/ 00,110"	-
テーマコード	String	"5A000","5A001"	-
Fターム	String	"AA01","AA02","AA03"	-
IPC	String	"A01A1/00","A01A1/01"	"A01A1/02","A01A1/03"
出願人名	String	特許太郎株式会社	Patent Inc.
発明者名	String	特許太郎	Patent Taro
更新日	Integer	20210101	20210101
テーマコード_機械	String	-	"5A002","5A003"
Fターム_機械	String	-	"AA04","AA05","AA06"
FI_機械	String	-	"A01A1/00,120","A01A1/ 00,130"
機械翻訳文_1	String	-	この発明は...
翻訳方式_1	String	-	SMT_model_1
機械翻訳文_2	String	-	この発明は...
翻訳方式_2	String	-	NMT_model_2
...	...	...	...
機械翻訳文_n	String	-	この発明は...
翻訳方式_n	String	-	NMT_model_n
図面のメタデータ	Binary	00010001,01110001..... (1),(1,0),(1,0,1,0),...	10011001,11110101..... (1),(1,0),(1,0,1,0),...
第1特徴ベクトル	Binary	(1),(1,0),(1,0,1,0),...	(1),(1,0),(1,0,1,0),...

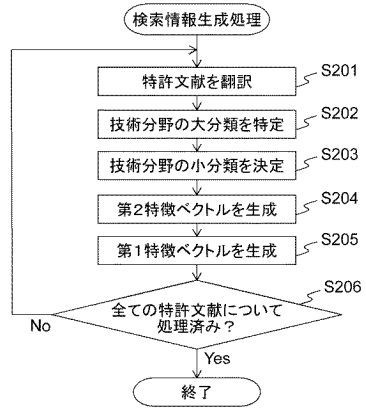
【 図 2 0 】

図20

フィールド名	データ型	レコード例 1	レコード例 2
発行機関	String	JPO	WIPO
文献番号	String	JP2020000000A	WO2020000000A
文献種別	String	A	A
格納場所	String	/var/www/bunken/data/text/JP/2020/000/000/	/var/www/bunken/data/text/WO/2020/000/000/
ファイル名	String	JP2020000000A.xml	WO2020000000A.xml
言語	String	ja	en
更新日	Integer	20210101	20210101
出願番号	String	JP2019000000	PCT/WO2019/000000
発行日	Integer	20200630	20200630
出願日	Integer	20190101	20190101
FI	String	"A01A1/00,100","A01A1/00,110"	-
テーマコード	String	"5A000","5A001"	-
Fターム	String	"AA01","AA02","AA03"	-
IPC	String	"A01A1/00","A01A1/01"	"A01A1/02","A01A1/03"
出願人名	String	特許太郎株式会社	Patent Inc.
発明者名	String	特許太郎	Patent Taro
更新日	Integer	20210101	20210101
テーマコード_機械	String	-	"5A002","5A003"
Fターム_機械	String	-	"AA04","AA05","AA06"
FI_機械	String	-	"A01A1/00,120","A01A1/00,130"
機械翻訳文_1	String	-	この発明は...
翻訳方式_1	String	-	SMT_model_1
機械翻訳文_2	String	-	この発明は...
翻訳方式_2	String	-	NMT_model_2
...	...	-	...
機械翻訳文_n	String	-	この発明は...
翻訳方式_n	String	-	NMT_model_n
図面のメタデータ	Binary	00010001,01110001.....	10011001,11110101.....
第1特徴ベクトル	Binary	{1},{1,0},{1,0,1,0},...	{1},{1,0},{1,0,1,0},...
テキストデータ	String	この発明は...	This invention is ...
二次データ	String	...	...
管理情報	String	...	...

【 図 2 1 】

図21



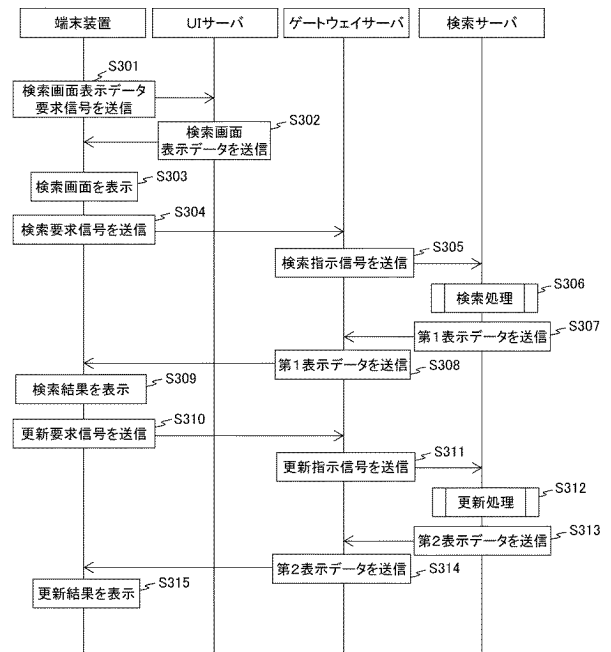
【 図 2 2 】

図22



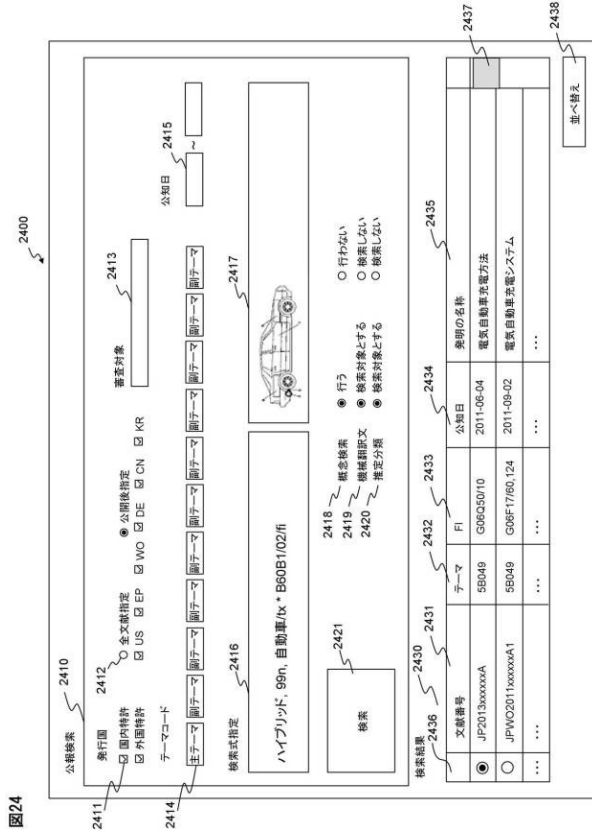
【 図 2 3 】

図23



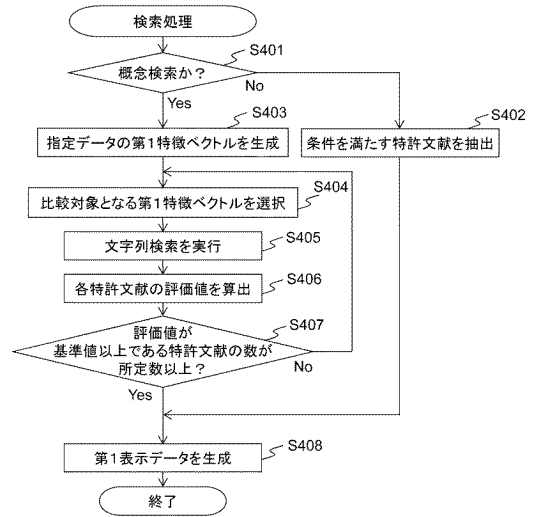


【図24】



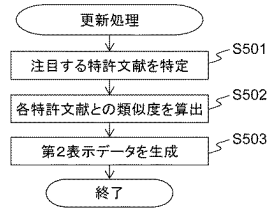
【図25】

図25



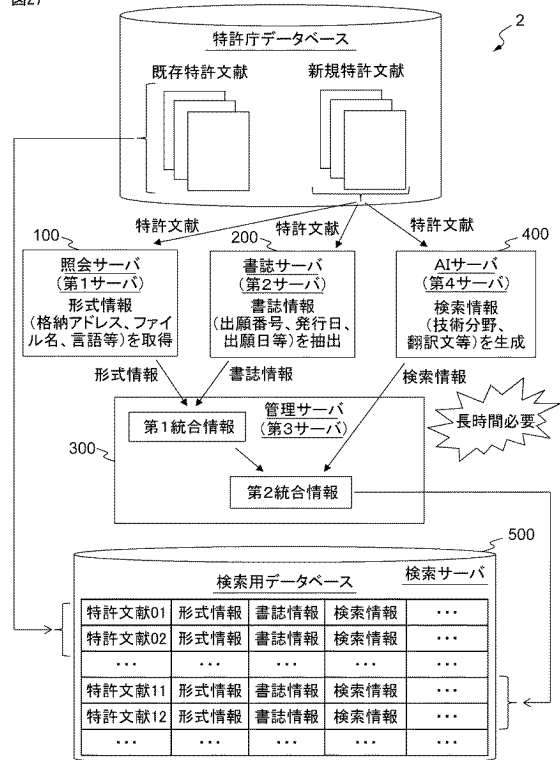
【図26】

図26



【図27】

図27



## フロントページの続き

- (72)発明者 久々宇 篤志  
東京都千代田区霞が関三丁目四番三号 特許庁内
- (72)発明者 後藤 昌夫  
東京都千代田区霞が関三丁目四番三号 特許庁内
- (72)発明者 関口 明紀  
東京都千代田区霞が関三丁目四番三号 特許庁内
- (72)発明者 西出 隆二  
東京都千代田区霞が関三丁目四番三号 特許庁内
- (72)発明者 間野 裕一  
東京都千代田区霞が関三丁目四番三号 特許庁内
- (72)発明者 目黒 光司  
東京都千代田区霞が関三丁目四番三号 特許庁内
- (72)発明者 森口 忠紀  
東京都千代田区霞が関三丁目四番三号 特許庁内

審査官 吉田 誠

- (56)参考文献 特開2014-119839(JP,A)  
特開2003-141168(JP,A)  
特開2007-199987(JP,A)  
特表2008-516341(JP,A)  
特開2009-211144(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 16/00 - 16/958